



**Politecnico  
di Torino**

**POLITECNICO DI TORINO**

Master of Science Degree in Civil Engineering

Master Thesis

**Clustering strategies in automated operational  
modal analysis for structural health monitoring**

Academic year 2020/2021

Supervisor:  
Professor Rosario Ceravolo

Candidate:  
Luigi Sibille

Associate Supervisor:  
Dr. Luca Zanotti Fragonara  
Dr. Marco Civera

CRANFIELD UNIVERSITY

Politecnico di Torino  
Dipartimento di Ingegneria Strutturale, Edile e Geotecnica

Master Thesis

Academic Year 2020 - 2021

Luigi Sibille

Clustering strategies in automated operational modal analysis for structural  
health monitoring

Supervisor: Professor Rosario Ceravolo  
Associate Supervisor: Dr. Luca Zanotti Fragonara  
Dr. Marco Civera

October 2021

© Cranfield University 2021. All rights reserved. No part of this publication  
may be reproduced without the written permission of the copyright owner.



# Abstract

In the last decades the interest in the dynamic properties of the structures soared rapidly in the field of Civil Engineering, due to the fast development of large-scale civil structures. The analysis of the dynamic behaviour can reveal progressive damages, which in turn allows to estimate the residual service life. Indeed, modal parameters and their evolution in time can be used as indicators of structural weaknesses or deficiencies induced by unforeseen events.

However, the impossibility to excite large size structures led to the use of output-only System Identification (SI) techniques, based on vibration data collected during working conditions. This framework is known as the Operational Modal Analysis. The main approach, utilised for this work as well, is the Stochastic Subspace Identification (SSI) method, applied for a range of model orders. However, this approach returns the candidate modes represented by means of a stabilization diagram, with many spurious modes, which need to be (manually or automatically) removed. This is the ambit of Automated Operational Modal Analysis (AOMA); despite the widespread interest from the scientific community, some challenges in improving its reliability and computational cost-effectiveness are still open. In this study, a novel two-stage clustering approach for AOMA is proposed. In the first stage, the cleansing out of certainty mathematical poles from the diagram is performed via K-means clustering algorithm, whereas the DBSCAN clustering is developed for detecting and identifying the remaining outliers. Contrarily to the existing approaches, DBSCAN clusters the poles and detect the outliers in only step. The parameters needed to perform the clustering algorithm are automatically estimated using a cluster validation criterion and heuristic methods. Several numerical and experimental case studies are here reported for the validation of the procedure, to illustrate the robustness and the performance of the proposed AOMA

method. The Z24 bridge benchmark, an experimental case study regarding a helicopter blade and a numerical case are indeed analyzed.

# Acknowledgement

I would like to thank my associate supervisor Dr. Luca Zanotti Fragonara for giving me patient support, instructive advice, and useful suggestions on my work. I also strongly appreciate, as despite his busy schedule and the organizational problems due to the pandemic, his availability to discuss any issues I had during the thesis. Additionally, I would like to express gratitude to Dr. Marco Civera for his constant support which was really influential in shaping the algorithm.

I'm extremely grateful to my supervisor Rosario Ceravolo for giving me the opportunity to carry out all my thesis work at Cranfield University enhancing my educational path.

Last but not least, I would like to thank my family who always believe me for the financial and mental support.

# List of Contents

<b>Abstract</b> .....	i
<b>Acknowledgement</b> .....	iii
<b>List of Contents</b> .....	iv
<b>List of Figures</b> .....	vii
<b>List of Tables</b> .....	ix
<b>1 Introduction</b> .....	1
<b>1.1 Problem statement</b> .....	1
<b>1.2 Aims and objectives</b> .....	3
<b>1.3 Thesis outline</b> .....	4
<b>2 Literature Review</b> .....	6
<b>2.1 Structural Health Monitoring</b> .....	6
<b>2.2 Operational Modal Analysis</b> .....	7
<b>3 Theoretical background</b> .....	9
<b>3.1 Comparison parameters</b> .....	9
<b>3.1.1 Frequency, damping ratio and eigenvalue</b> .....	11
<b>3.1.2 Modal Assurance Criterion</b> .....	12
<b>3.1.3 Mean Phase and Mean Phase deviation</b> .....	13
<b>3.2 Clustering algorithms</b> .....	13
<b>3.2.1 K-means clustering</b> .....	14
<b>3.2.2 Hierarchical clustering</b> .....	15
<b>3.2.3 DBSCAN</b> .....	16
<b>4 Methodology</b> .....	20
<b>4.1 Stochastic subspace identification method</b> .....	22
<b>4.2 Hard validation criteria</b> .....	28
<b>4.3 Soft validation criteria</b> .....	29
<b>4.3.1 Comparison parameters</b> .....	30
<b>4.3.2 Data pre-processing</b> .....	32
<b>4.3.2.1 Box-cox transformation</b> .....	32
<b>4.3.2.2 Standard score standardization</b> .....	32

4.3.3	K-means clustering.....	33
4.4	Cluster identification.....	35
4.4.1	DBSCAN.....	35
4.4.1.1	MinPts determination .....	37
4.4.1.1.1	Silhouette index .....	37
4.4.1.2	Epsilon determination .....	40
4.4.1.2.1	Elbow rule.....	40
4.4.2	Distance vector.....	41
4.5	Final modal parameter selection.....	42
4.6	Major differences with previous proposals.....	43
5	Numerical case .....	45
5.1	Description.....	45
5.2	Automated Operational Modal Analysis.....	49
5.2.1	Hard validation criteria.....	49
5.2.2	Soft Validation Criteria.....	51
5.2.3	DBSCAN.....	53
5.2.3.1	MinPts determination .....	56
5.2.3.2	Epsilon determination .....	56
5.2.4	Final modal parameters.....	57
5.3	Influence of signal duration and noise .....	64
5.3.1	Precision, Recall and F-score .....	64
5.3.2	Results .....	66
5.3.3	Conclusions.....	70
6	Helicopter blade.....	72
6.1	Description.....	72
6.2	Automated Operational Modal Analysis.....	73
6.2.1	DBSCAN.....	74
6.2.1.1	MinPts determination .....	76
6.2.1.2	Epsilon determination .....	76
6.2.1.3	Final Modal Parameters .....	77
6.3	Comparison with Mugnaini et al's results .....	80

<b>7</b>	<b>Z24 bridge</b> .....	<b>83</b>
7.1	<b>Description</b> .....	83
7.2	<b>Experimental setup</b> .....	84
7.3	<b>AOMA application and comparisons</b> .....	86
<b>8</b>	<b>Conclusions</b> .....	<b>97</b>
8.1	<b>Summary and conclusions</b> .....	97
8.2	<b>Future perspectives</b> .....	98
	<b>References</b> .....	<b>100</b>
	<b>APPENDICES</b> .....	<b>1</b>

# List of Figures

Figure 3-1 Comparison of clustering approaches .....	19
Figure 4-1 Overview of the linear time-independent model .....	25
Figure 5-1 Disposition of accelerometers .....	46
Figure 5-2 Stabilization diagram obtained before HVC .....	50
Figure 5-3 Stabilization diagram obtained after HVC .....	50
Figure 5-4 Stabilization diagram after SVC .....	52
Figure 5-5 Damping ratio-model order diagram after SVC.....	52
Figure 5-6 Damping ratio-model order-frequency diagram after SVC.....	53
Figure 5-7 Number of poles in each cluster .....	54
Figure 5-8 Silhouette values .....	54
Figure 5-9 Stabilization diagram after DBSCAN .....	55
Figure 5-10 Damping ratio-model order-frequency diagram after DBSCAN.....	55
Figure 5-11 MinPts determination.....	56
Figure 5-12 Epsilon determination .....	57
Figure 5-13 1 <sup>st</sup> mode shape identified .....	59
Figure 5-14 2 <sup>nd</sup> mode shape identified.....	59
Figure 5-15 3 <sup>rd</sup> mode shape identified .....	60
Figure 5-16 4 <sup>th</sup> mode shape identified .....	60
Figure 5-17 5 <sup>th</sup> mode shape identified .....	61
Figure 5-18 6 <sup>th</sup> mode shape identified .....	61
Figure 5-19 7 <sup>th</sup> mode shape identified .....	62
Figure 5-20 8 <sup>th</sup> mode shape identified .....	62
Figure 5-21 9 <sup>th</sup> mode shape identified .....	63
Figure 5-22 Precision in terms of difference of frequencies.....	67
Figure 5-23 Precision in terms of MAC.....	67
Figure 5-24 Recall in terms of difference of frequencies .....	68
Figure 5-25 Recall in terms of MAC .....	68
Figure 5-26 F-score in terms of difference of frequencies.....	69
Figure 5-27 F-score in terms of MAC.....	69

<b>Figure 6-1 (a) Airbus Helicopter H135 general arrangement. (b) main parts of a BMR blade. (c): picture of the investigated BMR blade, acquisition system, and experimental setup .....</b>	<b>73</b>
<b>Figure 6-2 Number of poles in each cluster .....</b>	<b>74</b>
<b>Figure 6-3 Silhouette index .....</b>	<b>75</b>
<b>Figure 6-4 Identified clusters after DBSCAN.....</b>	<b>75</b>
<b>Figure 6-5 MinPts determination.....</b>	<b>76</b>
<b>Figure 6-6 Epsilon determination .....</b>	<b>77</b>
<b>Figure 6-7 Silhouette value of the clusters identified by Mugnaini .....</b>	<b>80</b>
<b>Figure 6-8 3D stabilization diagram.....</b>	<b>81</b>
<b>Figure 6-9 Cluster identification .....</b>	<b>82</b>
<b>Figure 7-1 Z24 Bridge, Koppigen-Utzenstorf, Switzerland.....</b>	<b>83</b>
<b>Figure 7-2 Cross-section of the Z24-Bridge .....</b>	<b>84</b>
<b>Figure 7-3 Accelerometers location .....</b>	<b>85</b>
<b>Figure 7-4 Acquisition 300 2D clusters identification .....</b>	<b>87</b>
<b>Figure 7-5 Acquisition 300 3D clusters identification .....</b>	<b>87</b>
<b>Figure 7-6 Acquisition 605 2D clusters identification .....</b>	<b>88</b>
<b>Figure 7-7 Acquisition 605 3D clusters identification .....</b>	<b>88</b>
<b>Figure 7-8 Acquisition 1827 2D clusters identification .....</b>	<b>89</b>
<b>Figure 7-9 Acquisition 1827 3D clusters identification .....</b>	<b>89</b>
<b>Figure 7-10 Acquisition 3006 2D clusters identification.....</b>	<b>90</b>
<b>Figure 7-11 Acquisition 3006 3D clusters identification.....</b>	<b>90</b>
<b>Figure 7-12 Acquisition 4048 2D clusters identification.....</b>	<b>91</b>
<b>Figure 7-13 Acquisition 4048 3D clusters identification.....</b>	<b>91</b>
<b>Figure 7-14 The frequencies of the first 4 modes have been plotted as function of time by De Roeck from the first till the 250th day of monitoring and compared to the frequencies identified with the described AOMA algorithm. ....</b>	<b>92</b>
<b>Figure 7-15 Identified clusters of acquisition 4906.....</b>	<b>95</b>
<b>Figure 7-16 Stabilization diagram of acquisition 5595 after HVC .....</b>	<b>95</b>

# List of Tables

<b>Table 3-1 Advantages and disadvantages of three clustering algorithms.....</b>	<b>18</b>
<b>Table 5-1 Influence of each parameters on clustering process.....</b>	<b>51</b>
<b>Table 5-2 identification results of numerical case.....</b>	<b>58</b>
<b>Table 5-3 Conditions to determinate the type of mode shape .....</b>	<b>63</b>
<b>Table 5-4 Thresholds for acceptance as True Positive (TP) .....</b>	<b>65</b>
<b>Table 6-1 Frequencies identified for the experimental case study.....</b>	<b>77</b>
<b>Table 6-2 Damping ratios identified for the experimental case study .....</b>	<b>78</b>
<b>Table 6-3 Comparison of the identified frequencies with results of other studies ...</b>	<b>79</b>
<b>Table 7-1 Comparison of the identified frequencies with the Reynders' results.....</b>	<b>94</b>
<b>Table 7-2 Comparison of results obtained with different AOMA methods .....</b>	<b>96</b>

# 1 Introduction

## 1.1 Problem statement

The number of aged and damaged civil engineering infrastructure increased consistently in the last decades. Therefore, the observation and monitoring of changes of the material and geometrical properties over time became a new important area of research with the aim of maximizing the lifespan and security of structures. The interaction of people with old infrastructure is indeed becoming more and more frequent, therefore the sensibility of the public opinion on the importance of the maintenance of historical and old buildings raised as well.

The Structural Health Monitoring (SHM) involves the evaluation and monitoring of structural health. It has been applied in several engineering fields with the aim of improving structural reliability, increasing the lifespan of the structures, monitoring the damage propagation, and safeguarding human lives. The acquisition of parameters that characterized the behavior of the structure is performed through the use of different methods. This choice depends on the purposes of the monitoring, the economical budget, and the object of study. One of the most important analysis is the determination of the dynamical characteristic.

The pioneering works on modal testing are traced back to the 1940s, when engineers were charged with conducting vibration tests on aircraft structures with the attempt to understand when the structure would fail under dynamic loads. Only after the 1970s the investigations have been gradually extended to fields different than aerospace and mechanical engineering aiming to ensure that natural vibration frequencies were not the same as excitation frequencies, thus guaranteeing safety standards. In the last decades the interest in studying the dynamic properties of the structures in the field of civil engineering soared rapidly due to the fast development of large-scale civil structures and to ensure more safety avoiding mortal accidents

as the Mexico City earthquake in 1985 or the Tacoma Narrows bridge disaster in 1940.

Differently from mechanical and aerospace cases where the excitation and the response (input and output) records are usually known, in civil engineering it is not possible to excite with artificial devices high buildings or large span bridges, so usually only the output information is available to perform the modal analysis. Therefore, the dynamic properties of structures under working conditions are identified by using different methods such as the operational modal analysis (OMA).

The OMA methods consequently rely on the implementation of system identification methods for identifying the candidate modal parameters, one of the most reliable and efficient is the stochastic subspace identification (SSI) method. The SSI is one of the most popular modal identification algorithms in time domain which was presented by Van Overschee and Moor in 1991. It overcomes some typical flaws of the frequency-based methods as the bad inaccurate identification of closed modes that are typical of flexible structures and the insufficient resolution in frequency domain. Moreover, it seems to be less susceptible to the record length even though longer durations usually correspond to more accurate and reliable estimations [1]. The number of identified candidate modes is correlated to the selected model order that is usually defined as a range of model orders to overcome typical uncertainties. Therefore, the physical modes representative of the vibration behavior are distinguished for their repetition in different model orders. The identification is eased by the use of a tool that is called stabilization diagram. It is simply a plot of different model orders or damping ratios versus the frequencies correlated to each model order, thus the physical modes form identifiable vertical lines and can be easily distinguished. However, this process needs the human interaction that involves time and user experience. In recent years, validation criteria and clustering algorithms permitted the automation of this process defining thresholds and qualitative analysis. In the following chapters a novel application of the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) clustering algorithm for recognizing the

physical poles and interpreting the stabilization diagram is presented. The algorithm is partially based on the studies conducted by Reynders et al [2], Neu et al [3], and Mugnaini et al [4].

### **1.2 Aims and objectives**

The aim of the thesis is to implement a novel automated operational modal analysis (AOMA) method based on machine learning approaches. The cleansing out of the stabilization diagram identified by the SSI algorithm is completely automated and it does not need any user interaction. The definition of criteria to assess the similarity of the modes and to distinguish physical and mathematical modes are the challenges of this thesis. These achievements are accomplished through the following key points:

- The research of similar studies to be familiar with comparison parameters, validation criteria, clustering algorithms and basic statistical tools used to identify the final modal parameters
- The understanding of the theoretical background of the abovementioned instruments
- The implementation of a novel AOMA method based on the DBSCAN algorithm
- The automatization of the DBSCAN by means of a heuristic method and a clustering evaluation index
- The application of the proposed method on a numerical case
- The application of the proposed method on an experimental case that regards a helicopter blade
- The application of the proposed method on a real existed structure. The Z24 bridge is indeed investigated.

### 1.3 Thesis outline

The thesis is divided in the following chapters:

- **Chapter 1:** The thesis is briefly contextualized reporting the main aspects that led to study the operational modal analysis. The aims and objectives that drove this research are reported as well.
- **Chapter 2:** A critical review of the literature regarding the Structural Health Monitoring and operational modal analysis is provided. How the modal analysis of operating civil engineering infrastructures is performed and the differences with respect the experimental modal analysis are explained.
- **Chapter 3:** Several articles are examined in other to present a summary of the methods used and the basis of this work. This chapter is divided in the following sections:
  - Comparison parameters
  - Clustering algorithms
- **Chapter 4:** The relevant underpinning theory, the experimental techniques, and the explanation of the proposed automated OMA method are described in this chapter. The chapter is divided in the following steps:
  - Stochastic subspace identification method
  - Hard validation criteria
  - Soft validation criteria
  - Clusters identification
  - Final modal parameters selection
- **Chapter 5:** The AOMA method is firstly tested on a numerical case to assess the performance and robustness of the algorithm. The numerical simulation concerns a three-storeys multi-bay shear type frame with 9 degree of freedom which modal parameters obtained through the eigenproblem are compared to the modal parameters recognized with the AOMA algorithm. Besides, the

signal length as well as the level of noise artificially added to the signal are varied to evaluate their influence on the results.

- **Chapter 6:** The AOMA method is applied on an experimental case study regarding a helicopter blade. The data have been acquired during recent studies at Cranfield University on which the acceleration of flapwise and edgewise sensors has been recorded (nineteen channels in total). The modal parameters obtained with the AOMA have been compared to the ones benchmarked with the FEM and to the ones identified with the Mugnaini et al's algorithm [4].
- **Chapter 7:** The really existed Z24 bridge has been finally analyzed. The bridge has been subject of long-term continuous monitoring during the year before demolition in the framework of the SIMCES project. The AOMA is performed on different output signals and compared with the results of B. Peeters, De Roeck [5] and Reynders et al [2].
- **Chapter 8:** In the last chapter an overview of the thesis is explained with particular attention to the objectives achieved and to the suggestions on how the research could be taken by the following authors.

## 2 Literature Review

In this chapter the Structural Health Monitoring (SHM) is briefly introduced focusing on vibration-based techniques. Secondly the operational modal analysis is shortly exposed.

### 2.1 Structural Health Monitoring

To understand the behaviour of existing structures stressed with environmental or service loads, in the past few decades local material frameworks and performance evaluation methods applied on mathematical models have been employed. However, these methods rely only on simulation basis, not considering the uncertainty due to the real behaviour of structures, the construction phases, and the environmental noises. In order to overcome these problems, the continuous monitoring and analysis of civil infrastructures became a focus of many researchers. Moreover, the recent development and progress in sensor construction and information technology permitted the introduction of wave propagation-based and vibration-based techniques that aimed to study the global structural behaviour of the structures. All these techniques concern the Structural Health Monitoring (SHM) field.

The SHM method indeed entails the observation and monitoring of structures using periodically sampled response measurements over time diagnosing the evolution of the dynamic characteristics, updating mathematical models with field information, recognizing the beginning of damages and failures, and predicting any hazardous events that can lead not only to economic losses but also human lives risks. In other words, for long term the SHM provide information regarding the ability of the structure to perform its intend function giving that the aging and degradation are inevitable. Regarding extreme events, it is used for rapid screening and assessments of the integrity of the structure that can be followed by fast structural renovation.

Wave propagation-based SHM techniques are primarily used for local investigations where the wave lengths should be smaller than the dimension of the defects to be discovered. Local methods are consequently used if the position of the damage is known and therefore the sensors are localized in dense regions. Contrarily, in vibration-based SHM techniques the sensors are accelerometers that form a sensor network distributed over the whole structure. Quantitative analysis of any structure can be performed. Local sensors result as more accurate and sensitive to detect the damage, but the cost of sensor instrumentations is more expensive.

The aim of SHM is not only to detect any possible damage, but also monitor the performance of the structure. One of the main applications of SHM is indeed the modal analysis. Pioneering works of modal parameter identification are tracked back to the 1940s when the engineers began to understand the vibration behaviour of steel frameworks used in mechanical and aerospace field. Since then, with the development of sensors and computers the accuracy and the reliability of the modal analysis increased rapidly. The modal parameters identification in controlled environmental conditions, usually performed in laboratories is addressed by experimental modal analysis (EMA) methods. The EMA indeed uses only input-output vibration-based measurements derived from controlled excitation applied on the structure or framework of interest. However, the strength of EMA limits also its applicability for the impossibility to measure all the input forces in most of civil engineering structures.

Therefore, in the last decades other methods have been introduced for performing the modal analysis in massive structures such as the Operational Modal Analysis.

### **2.2 Operational Modal Analysis**

Operational modal analysis (OMA), also known as ambient modal identification, is a technique used to identify the modal properties of structures under its operating conditions. Cases where it's impossible to measure the ambient excitations, thus output only measurements are available, cannot be monitored and analyzed through

the use of experimental modal analysis but other techniques as the operational modal analysis are used. This method assumes that the input is a realization of stochastic processes also called white noise. However, a well-known problem of operational modal analysis methods is that if the white noise assumption is not satisfied, and the input contains also dominant frequency components, these frequencies cannot be separated during the processing and can be identified as noise as well.

The estimation of the modal parameters using the OMA can be summarized in three main steps: data collection, system identification, final modal parameters selection. The output only signal is collected as already explained through accelerometers positioned along the whole structure that monitor for long time. Among the robust and reliable system identification methods, one of the most used and popular is the stochastic identification method (SSI). The number of modes identified is directly correlated to one parameter specified by the user: the model order. Given that the signal and the process contain many biases the model order is overestimated, and several spurious modes are identified. To overcome this problem, the concept of stabilization diagram is introduced and the physical modes representative of the vibration behaviour of the structure can be easily identified considering their repetition in different model orders.

The automated OMA presented in the following chapter aims to automatize the selection of physical modes in the stabilization diagram. It doesn't require any user-defined parameter and can be applied to any SSI identification.

## 3 Theoretical background

In the last years, several studies have been carried out on the research of techniques or criteria designed to improve the efficiency and the performance of AOMA methods.

In this chapter, an excursus with brief theoretical contents is presented about the parameters, the machine learning techniques and the statistical tools implemented in the proposed AOMA method. Furthermore, the most used clustering algorithms are described, and their advantages and disadvantages are explained.

Firstly, the comparison parameters are explained and discussed. Secondly, the k-means, the hierarchical and the DBSCAN clustering algorithms are illustrated with a brief introduction of their more suitable usage situations.

### 3.1 Comparison parameters

Since the operational modal analysis relies on the evaluation of the similarity between different modes, the choose of the comparison parameters is a key step that has a direct influence on final results. Even if different approaches have been proposed over the years, all of them rely on the comparison of the modal variables, which can be summarized in:

- Frequency
- Damping ratio
- Modal shape

The first two modal variables are obtained from the identification of the eigenvalue that can be employed as a further modal variable to characterize the mode. It is worthy of note that the absolute difference or the relative difference among modal variables can be used to represent the variation of a system with different outlines, a short explanation of the reasons that lead to the use of the absolute difference is argued in the following paragraph. The stability of these modal variables is computed

between quantities that characterize modes belonging to following model orders. The comparison can be performed between two consecutive model orders (thus,  $n$  and  $n+2$ ) or two general following model orders (thus,  $n$  and  $n+2k$ ).

Differently, a multitude of approaches have been proposed over the years to compare the modal shape of the modes. With regards to the AOMA method proposed in this work, the complementary of the Modal Assurance Criterion (MAC) and the absolute difference of the Mean Phase Deviation have been implemented. Anyway, many other validation criteria can be easily found in the literature. Reynders et al [2] suggest to build a feature vector with as many validation criteria as possible, but on the other hand Neu et al [3] recommend “that more care must be taken to properly select, transform and normalize the variables for the feature vector”. The latter assertion refers to the incorrect identification of physical modes as mathematical due to the inappropriate use of validation criteria. Some criteria that measure the complexity and the excitation of mode shapes as the Modal Transfer Norm (MTN) and the Mean Phase Deviation (MPD) can lead the clustering algorithm to divide the modes into weak/strong or real/complex instead of evaluating their similarity. Moreover, further studies reported in [3] assert that the clustering computed by means of k-means algorithm is influenced by the probability distribution of the variables. Hence, variables with larger variances prevail in the clustering and consequently the physical poles are not identified properly. The proposed comparison parameters are a successful compromise of robustness and compatibility to different type of dataset primarily because they are based on the differences of modal variables. Some of the various validation criteria proposed in the literature like the Mode Overcomplexity Value (MOV) [6] and the Modal Scale Factor (MSF) [7] are not suitable for the OMA due to the lack of knowledge about the driving force or the mass-normalized mode shapes.

### 3.1.1 Frequency, damping ratio and eigenvalue

The difference in terms of frequency, damping ratio and eigenvalue between two modes belonging to two different model orders can be calculated through the use of absolute or relative differences.

The different approaches can be mathematically defined as follows:

- Absolute difference: 
$$\Delta X = |X_i - X_j|$$

- Relative difference: 
$$\Delta X = \frac{|X_i - X_j|}{\max(X_i - X_j)}$$

where  $X$  is the considered variable,  $i$  and  $j$  are the corresponding model orders.

The advantage of using the relative difference instead of the absolute one is that the obtained comparison parameters are already scaled from 0 to 1.

This normalization is particularly useful during the use of k-means clustering algorithm for cleansing the stabilization diagram. The parameters are indeed dimensionless and especially they have the same weight on the clustering.

Complete similarity between two modes is represented by a comparison parameter equal to 0, differently the value 1 means a perfect dissimilarity. These reasons drive many authors to choose the relative difference as the reference parameter.

However, referring to Mugnaini et al [4], the use of relative difference can lead to an underestimation of entities represented by low values for the following considerations.

Let the values  $a_1, a_2, b_1, b_2$  belong to the same generic variable and be the absolute distances between  $a_1, a_2$  and  $b_1, b_2$  equal. Mathematically it can be expressed as:

$$\Delta a = \Delta b = |a_1 - a_2| = |b_1 - b_2|$$

Supposing that  $a_1, a_2$  present values much smaller than  $b_1, b_2$ , it's clear that the absolute difference is different to the relative difference and consequently,

for  $a_1, a_2 \ll b_1, b_2$  and  $\Delta a = \Delta b = |a_1 - a_2| = |b_1 - b_2|$

$$\Delta a_r = \frac{|a_1 - a_2|}{\max(a_1, a_2)} \gg \Delta b_r = \frac{|b_1 - b_2|}{\max(b_1, b_2)}$$

If a threshold on the maximum relative difference among entities of a variable is imposed, the smaller values may be discarded even if representative of similarity.

As a matter of fact, the comparison parameters used in the presented AOMA algorithm are calculated as absolute difference and the problem of assigning the same scale to each variable is overcome using the standard score normalization. Differently, the problem of giving the same weight to the variables is addressed by means of the Box-cox transformation.

The above-mentioned techniques are necessary in order to properly perform the k-means clustering which considers different variables in the same process. All the methods are explained in the Methodology chapter.

### 3.1.2 Modal Assurance Criterion

The dimensionless statistical indicator Modal Assurance Criterion (MAC) is a very popular parameter that returns the degree of consistency of two different mode shapes which can be either real or complex [8]. It's sensitive to large differences and moderately insensitive to small differences. Moreover, it varies from 0 that indicates inconsistency between mode shapes, to 1 that means fully consistent mode shapes. Since the other parameters means equivalence for 0 values, the complementary of the Modal Assurance Criterion is used for easier understanding (0 for fully correlation, 1 for no correlation).

The Modal Assurance Criterion is defined as:

$$MAC(\phi_j, \phi_k) = \frac{[(\phi_j)^*(\phi_k)]^2}{[(\phi_j)^*(\phi_j)] * [(\phi_k)^*(\phi_k)]}$$

### 3.1.3 Mean Phase and Mean Phase deviation

Primarily the Mean Phase index is not computed by the AOMA algorithm, but it is provided by means of the system identification method SSI. The *MPD* indicates the mode shape complexity. It is the statistical variance of the phase angles for each mode shape coefficient from their mean value. Hence, it indicates the scatter of a mode shape about the mean phase angle. Its value tends to 0 for real normal mode shapes. Larger values than 0 means that the vector is complex mode oriented about the mean phase angle. For further details, refer to [9]. For the  $j^{th}$  mode shape, *MPD* and *MP* can be defined as:

$$MPD_j = \sqrt{\frac{\sum_{n=1}^N |\phi_{nj}| (\alpha_{nj} - MP_j)^2}{\sum_{n=1}^N |\phi_{nj}|}}$$

$$MP_j = \frac{\sum_{n=1}^N |\phi_{nj}| \alpha_{nj}}{\sum_{n=1}^N |\phi_{nj}|}$$

where  $\phi_{nj}$  is the  $n^{th}$  element of the  $j^{th}$  mode shape,  $\alpha_{nj}$  is its phase angle,  $N$  indicates the degree of freedoms and  $MP_j$  is the mean phase of the  $j^{th}$  mode shape.

## 3.2 Clustering algorithms

Generally, the automation of the interpretation of stabilization diagrams can be divided in two main steps: the cleansing from certainty mathematical poles or from outliers and the clustering of physical poles with the aim to identify the final modes representative of the system. Both the steps involve the implementation of clustering algorithms that have been proposed over the years with different purposes.

In this chapter the k-means, the (agglomerative) hierarchical and the DBSCAN clustering, which are three of the most popular clustering algorithms, are explained

with brief theoretical basis and applicative principles. In the last paragraph they are compared, and the advantages and disadvantages are described.

The first two techniques are very popular in automated OMA, in this respect, it's worth mentioning the AOMA methods proposed by Reynders et al [2], Neu et al. [3] and Mugnaini et al [4]. Contrarily, the DBSCAN clustering technique is a novel implementation in AOMA.

### **3.2.1 K-means clustering**

The k-means clustering, or Lloyd's algorithm is one of the most popular unsupervised machine learning algorithms that aims to partition the dataset into  $k$  predefined distinct groups (clusters) in which each observation belongs to only one cluster. The algorithm was first proposed in the field of signal processing in 1957 by Stuart Lloyd [10], even though it was published as journal article for the first time in 1982. Many k-means versions have been published over the years, three widely used techniques are the Forgy/Lloyd algorithm (1957, published 1982), the MacQueen algorithm (1967) and the Hartigan & Wong algorithm (1979). The choice of the best technique depends on the characteristic and on the dimension of the dataset and there is no best algorithm at all. Jain, Duin and Mao [11] suggest performing different techniques to have the best understanding of the dataset to select the more appropriate one. The algorithm used in the presented AOMA method is the Forgy/Lloyd k-means that is described in this paragraph and in the Methodology chapter. Basically, it assigns each observation to the closest cluster calculating the observation-centroid cluster distance. The first step to perform such algorithm is define the number of clusters  $k$  that divide the dataset. This selection can be user-defined or based on empirical methods such as the elbow rule method, the information criterion approach, the information theoretic approach, the cross-validation or performing the silhouette index [12]. Each approach presents different shortcomings depending on the dataset. Consequently, the  $k$  centroids of the clusters are identified by using  $k$  observations that are as much as possible far away from each other or assuming

random variables within the data space. The next step regards the association of all the observations to the nearest centroids relying on their distance. Different metrics can be used to quantify the distance. After the observations and clusters are coupled, the  $k$  centroids are re-calculated using the mean of the observations that belong to each cluster. This iteration is repeated until the centroids don't change and don't move anymore.

As previously mentioned, the variables processed by the k-means clustering must be scaled to give them the same weight and transformed with the aim of obtaining a distribution as much normal as possible.

In this work the variables are standardized through the use of the z-score standardization and normalized with the Box-cox transformation. The latter method is used also by Neu et al [3].

### **3.2.2 Hierarchical clustering**

Hierarchical clustering is one of most popular clustering algorithms in literature first proposed by S.C. Johnson in 1967 [13]. Contrarily to partitional clustering, which directly groups the poles into a set of disjoint clusters, the hierarchical clustering is a method that aims to build a hierarchy of clusters which can be graphically represented by a dendrogram. In order to create a partition of the data and thus determine the clusters, the dendrogram is cut at a certain level and all the objects below each cut are assigned to a single cluster.

Hierarchical clustering is generally classified into two types:

- Agglomerative clustering (bottom-up method): the clustering starts considering each observation as a different cluster. The clusters are then merged to the closest pair of clusters in respect of different linkage methods until the dendrogram is formed and all the clusters are grouped.
- Divisive clustering (top-down method): At the beginning all the observations are grouped in only one cluster and then are divided recursively as one moves

down the hierarchy until one singleton cluster remain. However, divisive clustering algorithms are not frequently used in practice.

There are many definitions of the distance and of the linkage between clusters, the most used are the single linkage, complete linkage, average linkage, Ward's linkage [14].

### 3.2.3 DBSCAN

In this chapter only a brief introduction of the Density-Based Spatial Clustering of Applications with Noise clustering algorithm is addressed considering that further explanations of the methodology used to perform this technique are reported in **Methodology** in **Cluster identification** paragraph.

DBSCAN is the first density-based clustering algorithm introduced for by Ester [15] in 1996. The theory on the basis is that a cluster in a data space is a contiguous dense region of points surrounded by points that lie in low density regions which are labelled as noise or outliers. Indeed, all the dense regions are instead identified and considered as clusters.

As most of the clustering algorithms, DBSCAN requires the definition of some parameters which are:

- Epsilon: the radius of the neighborhood around a data point  $p$
- MinPts: the minimum number of data points in the neighborhood to define a cluster

### Comparisons

Hundreds of clustering algorithms have been developed over the years for a wide variety of applications. Most of the existing cluster methods can be distinguished in the following classes:

- Partitioning clustering methods: Identify clusters as highly populated data dividing the dataset in regions with similar dimension.

- Hierarchical clustering methods: Provide a nested sequence of clusters organized as a hierarchical tree.
- Density based clustering methods: Cluster dense connected regions of data which can have a flexible shape.

Moreover, they can be classified in further subclasses. Hard partitioning clustering assigns the object to only one cluster, instead in the fuzzy partitioning clustering the object is assigned to the cluster with a degree of membership ranging between 0 to 1. As mentioned before, also the hierarchical clustering can be divided in two subclasses which are the agglomerative and the divisive. Referring to the clustering techniques introduced in the previous paragraphs, the k-Means pertains to the fuzzy partitioning class, the DBSCAN to the density-based clustering and the hierarchical to the hierarchical clustering.

Besides, partitional and density-based algorithms usually require stronger assumptions such as the number of clusters and the initial centers for the k-means or the epsilon and MinPts for the DBSCAN to be executed but they are faster than the hierarchical algorithms.

Advantages and disadvantages of the introduced methods are summarized in

*Table 3-1:*

## Theoretical background

---

	<b>Advantages</b>	<b>Disadvantages</b>
<b>K-means</b>	<ul style="list-style-type: none"><li>• Easy to implement</li><li>• Computationally efficient</li></ul>	<ul style="list-style-type: none"><li>• Difficult to predict the number of clusters</li><li>• Very sensitive to normalization or standardization of the data</li></ul>
<b>DBSCAN</b>	<ul style="list-style-type: none"><li>• Flexibility in shape and size of the clusters</li><li>• Detects outliers</li></ul>	<ul style="list-style-type: none"><li>• Doesn't work well with data with various density</li></ul>
<b>Hierarchical</b>	<ul style="list-style-type: none"><li>• Easy to implement</li><li>• Provides informative hierarchical interpretation</li></ul>	<ul style="list-style-type: none"><li>• Very sensitive to noise</li><li>• Not computational efficient</li><li>• Rarely provides the best solution</li></ul>

*Table 3-1 Advantages and disadvantages of three clustering algorithms*

A comparison of samples of different datasets processed with the proposed three clustering algorithms and their time of execution are in *Figure 3-1*.

## Theoretical background

---

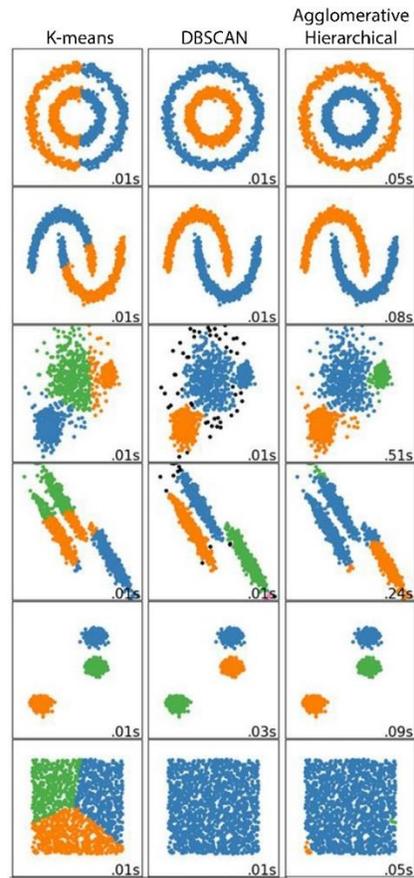


Figure 3-1 Comparison of clustering approaches

In this work, the capability of detecting outliers and better clustering connected dense regions of points, leads the Authors to use the DBSCAN instead of the k-means and the hierarchical.

## 4 Methodology

In Structural Health Monitoring the Operational Modal Analysis (OMA) is one of the most popular and efficient techniques for identifying the modal parameters of a structure based on vibration data collected during operating conditions. The proposed automated OMA (AOMA) aims to automatically evaluate the candidate modal parameters distinguishing between mathematical (spurious) and physical modes without manual user interaction. The discrimination is carried out comparing the modal parameters identified by means of the SSI system identification method relying on the empirical observations that the physical modes are almost identical at every system order. Contrarily, the mathematical modes occur for the presence of noise or for the over-estimation of the system order in the SSI. As reported by C. Rainieri and G. Fabbrocino [1] “Typical noise sources are modelling inaccuracies (for example, the system that generated the data cannot be modelled exactly as a stochastic state-space model), measurement noise (due to sensors and measurement hardware), computational noise (due to the finite precision of computers), the finite number of data points (as a consequence, the factorization property of the Toeplitz or projection matrix does not hold exactly and their rank will not exactly be  $n$ )”. The challenge is addressed by the use of validation criteria and machine learning approaches. The stabilization diagram is cleansed through hard validation criteria (threshold values) and soft validation criteria (based on the similarity between modes and on a clustering algorithm) that evaluate the physical meaning of the modes. Thereafter, the possible physical modes are clustered through the DBSCAN algorithm that automatically identify the spurious modes as noise. Differently from the usual implementation of the DBSCAN that requires the specification of two parameters, in the presented AOMA algorithm the Authors propose an automated parameter estimation using a cluster validation criterion and a very popular heuristic method in clustering analysis. Lastly, the final modal

representation of each cluster is estimated applying five different statistical approaches.

The AOMA procedure can be described in five steps:

### **1. System identification method**

The candidate modes are identified by the Stochastic Subspace Identification methods. Since the SSI don't distinguish the physical and the mathematical modes further steps are necessary. The selection of the required parameters to perform the SSI are described in the following paragraphs.

### **2. Hard validation criteria**

The HVC is applied to all the candidate modes found by the SSI. The modes that do not present a physical meaning are labelled as certainty mathematical and discarded by the use of fixed threshold values.

### **3. Soft validation criteria**

This step aims to evaluate the consistency of the modes at different system orders. The poles that show similarity are considered as possible physical otherwise they are flagged as spurious, and they are not considered in the following steps. The consistency is measured computing the differences between eigenfrequencies, damping ratios, eigenvalues and comparing two parameters which regard the mode shapes. The clustering is performed by means of the k-means clustering algorithm.

### **4. Density-Based Spatial Clustering of Applications with Noise clustering**

The main purpose of this step is to automatically detect the stabilization axes in the cleansed stabilization diagram precisely and clearly. The DBSCAN clusters the physical poles with similar modal parameters and recognizes the spurious poles as outliers in only one step. In contrast to the usual use of the DBSCAN, the clustering algorithm doesn't need any user-defined parameters. Moreover, a cluster validation criterion is performed and plotted to show the quality of the clustering.

## **5. Final modal parameter selection**

Since each cluster contains a large quantity of poles, the modal parameters representative of each cluster are computed by using five different methods that are based on basic statistical approaches.

### **4.1 Stochastic subspace identification method**

Every AOMA algorithm analyzes a set of poles that are firstly identified by means of a system identification method. In the time domain one of the most advanced and efficient modal identification techniques is the covariance driven stochastic subspace identification (SSI-COV) [16]. The SSI-COV identifies the modal parameters from the state space matrix relying on output correlations [17]. Hence, differently from the data-driven SSI (SSI-data) that directly deals with raw data, the SSI-COV indeed works with output-only records.

Although the SSI permits an effective and robust identification even with not stationary data, not all the identification results are representative of physical modes. The identified number of poles are directly correlated to the model order that has to be chosen high enough to allow the identification of all the physical modes and sufficiently low to prevent the over-modelling that lead to the identification of spurious poles. For the model order  $n$ , SSI identifies  $n$  sets of modal parameters composed of  $n$  eigenvalues and  $n$  mode shapes. Since one model order  $n$  is not enough to identify which are the physical modes of the system, in order to overcome this uncertainty a range of model orders are commonly specified. Since the physical poles remain quite stable varying the model order, differently to the spurious poles that vary for the presence of noise, they are recognized for their repetition in different model orders using the stabilization diagram. Subsequently the physical poles are identified from the stabilization diagram through the use of data clustering algorithms as explained in the following paragraphs. In the presented research, the range of the model order varies on a case-by-case basis similarly to the studies carried out by

Mugnaini et al. Therefore, the model order is increased by two ranging from  $n_{min}$  and  $n_{max}$ , i.e.  $n_{min}, n_{min} + 2, n_{min} + 4, \dots, n_{max}$ .

Actually, the application of SSI involves the definition of another parameter that influences the accuracy of the modal parameter estimations. The role of the block rows of the Hankel matrix has been deeply investigated as also how an incorrect setting can lead to biased results [1]. It is a well-established fact that according to Reynders and De Roeck [18] the block rows of the Hankel matrix should respect the following rule of thumb:

$$i \geq \frac{f_s}{2 f_0} \quad (4-1)$$

where  $f_s$  is the sampling frequency and  $f_0$  is the lowest frequency of interest.

In order to always verify the above-mentioned criterion in all the reported case studies the block rows of the Hankel matrix has been set as follow:

$$i = \frac{f_s}{2 f_0} \quad (4-2)$$

The stochastic subspace identification method is introduced with brief theory content below.

At present, two implementations of the SSI have been presented: the data driven SSI (SSI-DATA) and the covariance-driven SSI (COV-SSI). The data driven directly deals with raw output data, whereas the covariance driven works with output-only records. Even if both the methods have similar results, the uncertainties of the modal parameters are easily found with the SSI-COV. SSI entails several robust numerical techniques in order to improve the cost efficiency of calculations such as the least-square fitting, the SVD and the QR decomposition. Although great achievements regarding the modal analysis based on output-only information, the SSI and generally the OMA procedures are still worth improving the accuracy of the identification and reducing the computational cost of the algorithms. Studies on the influence of analysis parameters (model order and Henkel matrix) conducted by

Ranieri and Fabbrocino on the computational time and the accuracy of modal parameter estimates are not reported for conciseness. The SSI algorithm described in this section is the covariance-driven which is used in the presented AOMA algorithm. It can be summarized in the following steps:

1. Identification of the stochastic state-space model
2. Computation of output covariance
3. Construction of the Toeplitz matrix
4. Decomposition of the Toeplitz matrix
5. Estimation of controllability and observability matrices
6. Estimate of the modal parameters.

Referring to Van Overschee and Moor [19], the systems are represented by a linear time-invariant stochastic model which is depicted in *Figure 4-1*. The circled vectors  $u_k$  and  $y_k$  are determinist signals measured at the time  $k$  of respectively the  $m$  input and the  $l$  output of the process, whereas  $w_k$  and  $v_k$  are unknown disturbances independent of  $u_k$  and  $y_k$  referred to the noise due to external factors or measurement inaccuracies. The vector  $x_k \in R^n$  is the state vector measured at the time  $k$  which contains the numerical state of  $n$  values, the matrix  $B \in R^{n \times m}$  is the input matrix, the matrix  $A \in R^{n \times n}$  is the (dynamic) system matrix and describes the dynamics of the system,  $\Delta$  represents a delay, the matrix  $C \in R^{l \times n}$  is the output matrix that describes how the internal system is transferred to the output vector  $y_k$  and the matrix  $D \in R^{l \times m}$  is called the direct feedthrough term.

Mathematically the model can be express by the two following difference equations:

$$x_{k+1} = Ax_k + Bu_k + w_k \quad (4-3)$$

$$y_k = Cx_k + Du_k + v_k \quad (4-4)$$

The ambient noise  $w_k$  and  $v_k$  are assumed to be stationary, white noise and zero mean. Their covariance matrix is defined as:

$$E \begin{bmatrix} w_p \\ v_p \end{bmatrix} \begin{bmatrix} w_q^T & v_q^T \end{bmatrix} = \begin{pmatrix} Q & S \\ S^T & R \end{pmatrix} \delta_{pq} \quad (4-5)$$

$$Q = E[w_p \ w_p^T], S = E[w_p \ v_p^T], R = E[v_p \ v_p^T] \quad (4-6)$$

where  $E$  denotes the mathematical expectation operator,  $\delta_{pq}$  is the Kronecker operator,  $p$  and  $q$  represents two time instant.

Since the input signal  $u_k$  is conventionally considered to be stationary zero mean Gaussian white noise as mentioned before, the eq. (4-3) and (4-4) can be reduced to:

$$x_{k+1} = Ax_k + w_k \quad (4-7)$$

$$y_k = Cx_k + v_k \quad (4-8)$$

where  $E[w_k \ x_k] = 0$  and  $E[v_k \ x_k] = 0$ ,

Consequently, a recursive relationship between output covariances

$R_i = E[y_{k+1} \ y_k^T]$  and system state matrix  $A$  can be defined:

$$\begin{cases} R_i = CA^{i-1}G \\ R_{-i} = G^T(A^{i-1})^T C^T \end{cases} \quad (i = 1, 2, 3 \dots) \quad (4-9)$$

where  $G = E[x_{k+1} \ y_k^T]$  is the covariance matrix of the output and the next state.

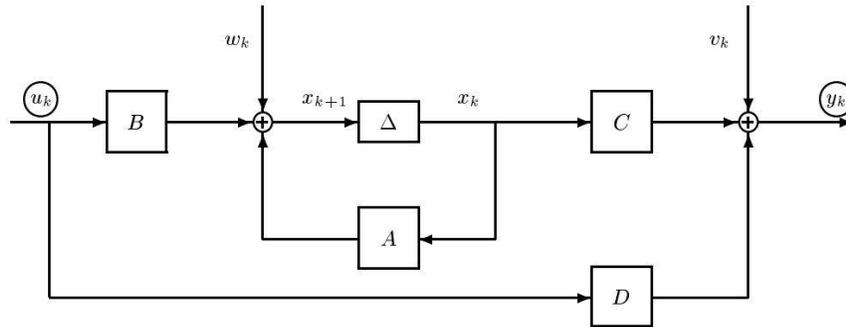


Figure 4-1 Overview of the linear time-independent model

For the computation of output covariance, firstly all the signals are gathered in the Henkel matrix  $H \in R^{2li \times j}$  with past and future time horizon defined as follow:

$$H = \frac{1}{\sqrt{j}} Y_{0/2i-1} = \frac{1}{\sqrt{j}} \begin{bmatrix} y_0 & y_1 & \dots & y_{j-1} \\ \dots & \dots & \dots & \dots \\ y_{i-2} & y_{i-1} & \dots & y_{i+j-3} \\ y_{i-1} & y_i & \dots & y_{i+j-2} \\ y_i & y_{i+1} & \dots & y_{i+j-1} \\ y_{i+1} & y_{i+1} & \dots & y_{i+j} \\ \dots & \dots & \dots & \dots \\ y_{2i-1} & y_{2i} & \dots & y_{2i+j-2} \end{bmatrix} = \frac{1}{\sqrt{j}} \begin{pmatrix} Y_{0/i-1} \\ Y_{i/2i-1} \end{pmatrix} = \frac{1}{\sqrt{j}} \begin{pmatrix} Y_p \\ Y_f \end{pmatrix} \quad (4-10)$$

The Hankel matrix  $Y_{0/2i-1}$  is defined by  $2i$  block rows and  $j$  columns. The number of rows  $i$  is defined by the user, whereas the number of columns  $j$  is usually assumed equally to the full-length of the signal. For each row  $i$ ,  $l$  sub-rows are defined, being  $l$  equals to the number of output channels ( $y_k \in R^{l \times 1}$ ). The Henkel matrix is divided in two sub matrices with  $i$  rows each representing the so-called future ( $Y_f \in R^{li \times j}$ ) and past ( $Y_p \in R^{li \times j}$ ) observation matrices.

Subsequently, the output covariance matrix is calculated as:  $\Lambda_i = E[y_{k+1} y_k^T]$ . By assuming the output signal obtained by an ergodic process, the output covariance matrix can be defined by using the associated estimator  $\hat{\Lambda}$ :

$$\hat{\Lambda} = \frac{1}{j} \sum_{k=0}^{j-1} y_{k+1} y_k^T \quad (4-11)$$

Consequently, the construction of the Toeplitz matrix follows. Through the use of the eq. (4-10) the Toeplitz matrix  $T_{1/i}$  is defined as follow:

$$T_{1/i} = Y_f Y_p^T = \begin{bmatrix} \Lambda_i & \Lambda_{i-1} & \dots & \Lambda_1 \\ \Lambda_{i+1} & \Lambda_i & \dots & \Lambda_2 \\ \dots & \dots & \dots & \dots \\ \Lambda_{2i-1} & \Lambda_{2i-2} & \dots & \Lambda_i \end{bmatrix} \quad (4-12)$$

Decreasing the dimension of the output data from the original form of the Hankel matrix with dimension  $2li \times j$  to the  $li \times li$  dimension of the Toeplitz matrix, the computational cost is significantly reduced.

Then the Toeplitz matrix is decomposed via singular value decomposition (SVD):

$$T_{1/i} = USV^T = [U_1 \quad U_2] \begin{bmatrix} S_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} = U_1 S_1 V_1^T \quad (4-13)$$

$$S_1 = \text{diag}[\sigma_i] \quad \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{n-1} \geq \sigma_n \quad (4-14)$$

where  $U$  and  $V$  are the orthogonal matrices with  $U_1, V_1 \in R^{l_i \times 2n_1}$  and  $S$  is the diagonal matrix with singular values with  $S_1 \in R^{2n_1 \times 2n_1}$ . The rank of the Toeplitz matrix indicates the order of structural system  $n$ .

The equation (4-13) can be rewritten as follows:

$$T_{1/i} = U_1 S_1 V_1^T = (U_1 S_1^{1/2} T) (T^{-1} S_1^{1/2} V_1^T) \quad (4-15)$$

where the non-singular matrix  $T$  has an order equal to  $n$ .

The estimation of controllability and observability matrices is computed rewriting the Toeplitz matrix:

$$T_{1/i} = \Gamma_i \Delta_i \quad (4-16)$$

$$\text{where } \Gamma_i = (C \quad CA \quad \dots \quad CA^{i-1})^T \text{ and } \Delta_i = (A^{i-1}G \quad \dots \quad AG \quad A) \quad (4-17)(4-18)$$

$\Gamma_i$  is the so-called extended observability matrix and  $\Delta_i$  is the reversed extended stochastic observability matrix.

If the matrix  $T$  is substituted by the identity matrix  $I$  in the eq. (4-15) and compared to the eq. (4-16), the eq. (4-17) and (4-18) can be identically written as:

$$\Gamma_i = U_1 S_1^{1/2} \quad (4-19)$$

$$\Delta_i = S_1^{1/2} V_1^T \quad (4-20)$$

For the estimation of the modal parameters, another Toeplitz matrix  $T_{2/i}$  can be calculated based on the eq. (4-13), (4-15), (4-19) and (4-20):

$$T_{2/i+1} = \Gamma_i A \Delta_i = \begin{bmatrix} \Lambda_{i+1} & \Lambda_i & \dots & \Lambda_2 \\ \Lambda_{i+2} & \Lambda_{i+1} & \dots & \Lambda_3 \\ \dots & \dots & \dots & \dots \\ \Lambda_{2i} & \Lambda_{2i-1} & \dots & \Lambda_{i+1} \end{bmatrix} \quad (4-21)$$

Thus, the Toeplitz matrix can be obtained by using only output records and the state matrix  $A$  and the output matrix  $C$  can be computed:

$$A = S_1^{-1/2} U_1^T T_{2/i+1} V_1 S_1^{-1/2} \quad (4-22)$$

$$C = \Gamma_i(1:l,:) \quad (4-23)$$

By means of eigenvalue decomposition the matrix A can be reformed as:

$$A = \Phi \Lambda \Phi^{-1} \quad (4-24)$$

where the diagonal matrix  $\Lambda$  contains the complex eigenvalues of the system  $\lambda_i$  and  $\Phi$  corresponds to the eigenvector matrix.

Finally, the frequencies, damping ratios and mode shapes can be identified by the following equations:

$$f_i = \frac{|\lambda_i|}{2\pi} \quad (4-25)$$

$$\varepsilon_i = \frac{-100\text{Re}(\lambda_i)}{|\lambda_i|} \quad (4-26)$$

$$\psi = C\Phi \quad (4-27)$$

## 4.2 Hard validation criteria

All the candidate modes identified by the SSI are represented in the stabilization diagram. The calculated frequencies are plotted in a graph with system orders as ordinate and frequencies as abscissa. The first cleanliness of the stabilization diagram that lead to the first identification of the certainty mathematical poles relies on the meaning of the modal parameters. Stable systems are not represented by negative damping ratios and moreover, regarding the cases studied in this work, the damping ratio of the poles cannot even be represented by values higher than 20%. The damping ratio of real systems not damped by special damping devices indeed definitely don't exceed the 20%.

Furthermore, real systems oscillate for the presence of complex conjugate pairs, therefore the eigenvalues  $\lambda_j$  must be composed by a positive real part and an imaginary part [3]. Hence, the first step of the presented AOMA algorithm concerns

the identification of some certainty mathematical poles imposing fixed thresholds. The just mentioned validation criteria can be mathematically represented as:

$$0\% \leq \xi_j \leq 20\% \quad (4-28)$$

$$Re(\lambda_j) \geq 0 \quad Im(\lambda_j) \neq 0 \quad (4-29)(4-30)$$

These criteria could be imposed at the beginning or at the end of the algorithm in the same way. The choice to impose the HVC as the first stage is made for reducing the computational cost of the algorithm.

The poles not identified as mathematical that pass the above-mentioned criteria are further evaluated in the next steps.

### 4.3 Soft validation criteria

Since many certain mathematical modes have been discarded in the previous section by means of fixed thresholds, in this part of the presented AOMA algorithm the modes are distinguished between possible physical and certainty mathematical on the basis of soft validation criteria. However, not all the certainty mathematical modes are identified in this stage but the remaining ones will be labeled as outliers in the following paragraph regarding the clustering algorithm.

The classification of the modes relies on the definition of five parameters used to quantify their similarity. A comparison five-dimension vector  $p_j^n$  is computed between the parameters of the pole  $j$  in the model order  $n$  and those of the most similar pole belonging to the following model order  $n+2$ .

Once the  $n$ -dimensional  $p_j^n$  vectors is computed for all the poles, the elements of the feature vector are standardized and transformed to obtain a distribution as much normal as possible.

The  $k$ -means clustering is finally performed to identify the more similar poles as stable.

### 4.3.1 Comparison parameters

Hence, these five comparison parameters concern the modal parameters of the poles. These are based on the eigenfrequencies  $f$ , the damping ratios  $\xi$ , the eigenvalues  $\lambda$  and two more parameters representing the mode shapes. They are defined as:

- The absolute difference between undamped eigenfrequencies  $d(f_j, f_k)$ :

$$d(f_j, f_k) = |f_j - f_k| \quad (4-31)$$

- The absolute difference between damping ratios  $d(\xi_j, \xi_k)$ :

$$d(\xi_j, \xi_k) = |\xi_j - \xi_k| \quad (4-32)$$

- The absolute difference between eigenvalues  $d(\lambda_j, \lambda_k)$ :

$$d(\lambda_j, \lambda_k) = |\lambda_j - \lambda_k| \quad (4-33)$$

this distance is redundant since combines eigenfrequencies and damping ratios as defined in the following formula:

$$\lambda_j = -|2\pi f_j| \xi_j + 2i\pi f_j \sqrt{1 - \xi_j^2} \quad (4-34)$$

- The dimensionless complementary of the *Modal Assurance Criterion* (MAC):

$$v_{j,k} = 1 - MAC(\phi_j, \phi_k) \quad (4-35)$$

where the Modal Assurance Criterion (MAC) is defined as:

$$MAC(\phi_j, \phi_k) = \frac{[(\phi_j)^*(\phi_k)]^2}{[(\phi_j)^*(\phi_j)]^*[(\phi_k)^*(\phi_k)]} \quad (4-36)$$

- The absolute difference of Mean Phase Deviation ( $\Delta MPD$ ):

$$\Delta MPD_{j,k} = |MPD(\phi_j) - MPD(\phi_k)| \quad (4-37)$$

The choice of using absolute differences instead of relative differences is due to the well-known issue that the relative differences can lead to an underestimation of the entities represented by low values. Consequently, in order to consider all the

variables in the same domain of existence the absolute differences are normalized through the z-score normalization.

Given that the distinction between modes is made evaluating their similarity to the others, modes with siblings in the following model orders have high possibility to be physical, otherwise they are certainly mathematical. Hence, for any  $j^{th}$  mode at the model order  $n$ , the five parameters above-mentioned are computed to evaluate their similarity to the closest mode in the following model order  $n+2$ .

Firstly, the five comparison parameters are computed for each mode in the model order  $n$  with respect to all the modes in the model order  $n+2$ . Considering  $d^n$  modes in the model order  $n$  and  $d^{n+2}$  modes in the model order  $n+2$ , each comparison parameter is computed  $d^n * d^{n+2}$  times to evaluate all the modes in the model order  $n$ .

The finding of the closest mode in the model order  $n+2$  is performed using the distance vector  $d_{j,k}^n$ . It is defined as follow:

$$d_{j,k}^n = \Delta f_{j,k} + v_{j,k} \quad (4-38)$$

where  $j$  is the  $j^{th}$  mode in the model order  $n$ ,  $k$  is the  $k^{th}$  mode in the model order  $n+2$ ,  $v_{j,k}$  is the complementary of the Modal Assurance Criterion and  $\Delta f_{j,k}$  is the min-max normalization of the absolute difference between undamped eigenfrequencies  $d(f_j, f_k)$  calculated as:

$$\Delta f_{j,k} = \frac{d(f_j, f_k) - \min(d(f_j, f_k))}{\max(d(f_j, f_k)) - \min(d(f_j, f_k))} \quad (4-39)$$

The min-max normalization is used to transform the range of  $d(f_j, f_k)$  into 0 to 1 for consistency with  $v_{j,k}$ . Therefore, the distance vector  $d_{j,k}^n$  is computed for each mode  $j$  in the model order  $n$  and its dimension is equal to the number of modes in the model order  $n+2$ . The mode  $k$  in the model order  $n+2$  that is closest to the mode  $j$  in the

model order  $n$  is the one that has the lowest distance, i.e., the lowest value in the distance vector  $d_{j,k}^n$ .

Once the nearest neighbor pole  $k$  in the following model order  $n+2$  has been found, the feature vector  $p_j^n$  calculated for any  $j$  mode in the model order  $n$  is computed with the above-mentioned validation criteria. It is reported as follows:

$$p_j^n = [d(\lambda_i, \lambda_k) \ d(f_j, f_k) \ d(\xi_i, \xi_k) \ v_{j,k} \ \Delta MPD_{j,k}] \quad (4-40)$$

### 4.3.2 Data pre-processing

As already explained, the k-means clustering performs better with variables with the same distribution belonging to the same domain of existence. The techniques used in the study are introduced in the following paragraphs.

#### 4.3.2.1 Box-cox transformation

Therefore, the power transformation is performed according to the Box-Cox transformation technique [20]. It's a process used to obtain a distribution as much normal as possible. The transformation from  $p_j^n$  to  $p(\lambda)$  is performed as follows:

$$p(\lambda) = \begin{cases} \frac{p-1}{\lambda} & \text{if } \lambda \neq 0 \\ \ln(p) & \text{if } \lambda = 0 \end{cases} \quad (4-40)$$

where  $p_j^n$  is the feature vector and the parameter  $\lambda$  is estimated maximizing the log-likelihood function. However, in this work the parameter  $\lambda$  has been automatically computed by the Matlab function `boxcox`.

This transformation must be executed for each parameter of all the poles.

#### 4.3.2.2 Standard score standardization

Since the transformed parameters are not in the same scale and therefore they have not the same influence on the clustering, the standard score standardization is performed. Hence, the standardized parameters have a mean equal to 0 and 1 as

standard deviation. The standard score standardization, also called the z-score normalization is defined as follows:

$$z_j^n = \frac{y_j^n - \bar{y}_p}{\sigma_p} \quad (4-41)$$

where  $\bar{y}$  is the average  $y$  parameter and  $\sigma$  is its standard deviation. These are calculated as:

$$\sigma_p = \sqrt{\frac{1}{\sum_{n=n_{min}}^{n_{max}} d^n} \sum_{n \in N} \sum_{j=1}^{d^n} (y_j^n - \bar{y}_p)^2} \quad (4-42)$$

$$\bar{y}_p = \frac{1}{\sum_{n=n_{min}}^{n_{max}} d^n} \sum_{n \in N} \sum_{j=1}^{d^n} y_j^n \quad (4-43)$$

consequently, the variables have the same influence on the clustering and the possible physical and certainty mathematical poles can be more correctly distinguished by the k-means.

### 4.3.3 K-means clustering

Briefly, the k-means clustering algorithm divides the poles in  $k$  clusters, each of which has its own centroid, relying on a predefined distance metric. The sum of the distances between every pole and the cluster centroids is minimized and the poles are assigned to its closest cluster. The aim is to flag the poles that don't show similarities (high values of the parameters) as certainty mathematical storing the possible physical ones (low values of the parameters) for the DBSCAN clustering algorithm.

The above-mentioned procedure can be summarized in the following steps:

1. Choose the number of clusters  $k$  and the position of their centroids
2. Compute observation-cluster centroid distances choosing the metric to use
3. Assign each observation to the cluster with the lower distance

4. Compute the mean of the observations in each cluster to obtain new  $k$  centroids
5. Repeat step 3 and 4 until the centroids don't change anymore.

Different metrics can be used to quantify the distance, the most common is the Euclidian distance that is calculated as follows:

$$dE = \sqrt{\sum_i^k (c_i - x_i)^2} \quad (4-44)$$

where  $x$  is the observation,  $c$  is the cluster centroid that is compared to,  $i$  is the dimension of  $x$  or  $c$  on which they are compared, and  $n$  is the total number of dimensions.

It is worthwhile to also mention the squared Euclidian distance:

$$dE^2 = \sqrt{\sum_i^k (c_i - x_i)^2} \quad (4-45)$$

the Manhattan distance:

$$dMht = \sum_i^k |c_i - x_i| \quad (4-46)$$

and the maximum distance between attributes of the vectors:

$$dMax = \max_{i=1\dots k} \sum_i^k |c_i - x_i| \quad (4-47)$$

In this dissertation the following assumptions have been implemented:

- The number of clusters is set equal to 2
- The centroids of the clusters are initially positioned in opposite site
- The chosen distance metric is the above-mentioned squared Euclidean distance
- The distances are computed between the feature parameters  $p$  of all the poles.

## 4.4 Cluster identification

After obtaining the cleared-out stabilization diagram as described in the preceding paragraphs, the physical and not-physical (computational) modes are distinguished by means of the density-based spatial clustering of applications with noise. This well-known data clustering groups points close to each other, marking the points that lie in low-density regions as outliers. The grouping is based on the similarity of the parameters. Referring to the previous work (Reynders, Neu and Vezio) [2][3][4], the DBSCAN aims to combine the hierarchical and the k-means clustering, identifying the clusters and removing the outliers in only one step. The main difference between DBSCAN and other clustering algorithms is indeed that not all the points form the final clusters but some are considered as noise. The case studies analyzed in the following chapters demonstrates the validity and the robustness of the algorithm.

### 4.4.1 DBSCAN

The performance of the DBSCAN is influenced by the following two parameters that generally are estimated manually accordantly with the distribution of points and the quantity of noise:

- *MinPts*: the minimum number of points necessary to form a dense region.
- *Eps*: the radius of the neighborhood around a pole. If the distance between two points is lower or equal to this threshold value, these points are considered as neighbor.

However, in this research, this selection is made automatically based on a cluster validation criterion and a very popular heuristic method in clustering analysis.

The key idea of DBSCAN is that for each pole the neighborhood calculated with a certain epsilon (eps) must contains a minimum number of poles (MinPts) to be considered as dense region, thus the neighborhood must overcome a predefined threshold.

Given the eps, the neighborhood of an arbitrary pole  $p$  is calculated as in [15]:

$$N_{eps} = \{q \in D / \text{dist}(p, q) < Eps\} \quad (4-48)$$

where  $D$  is the database of poles and  $q$  is another arbitrary pole.

If the pole  $p$  has a  $\varepsilon$ -neighborhood that contains at least the minimum number of MinPts points, this pole  $p$  is called core point. The definition of a core point is the following:

$$N_{eps}(P) > \text{MinPts} \quad (4-49)$$

On the contrary, if the condition is not satisfied the point  $p$  is considered as not-core point.

The points are considered as core points, reachable points and outliers based on the following definitions:

- A *core point* is a point that is surrounded by at least MinPts points within his  $\varepsilon$ -neighborhood
- A point  $q$  is *directly reachable* from  $p$  if it is within distance  $eps$  from the core point  $p$
- A point  $q$  is *reachable* from  $p$  if there is a path  $p_1, p_2, \dots, p_n$ , where  $p_1=p$ ,  $p_n=q$  and  $p_{i+1}$  is directly reachable from  $p_i$ . In this case all the points involved in the path must be core points except for  $q$ .
- All the points that are not reachable from any other point are labelled as *noise points* or *outliers*

The DBSCAN algorithm can be summarized in the following steps:

1. From the input data an arbitrary point  $p$  is chosen
2. the set of points directly reachable from  $p$  with respect to  $eps$  and MinPts is found
3. If the number of neighbors is higher than MinPts, the point  $p$  is labelled as core point and the first cluster  $C$  is initialized, otherwise the point  $p$  is considered as outlier or possible boarder point if later reached from core points. In both latter cases the algorithm returns to the step 1.

4. If the cluster C is initialized, the DBSCAN iterates over each neighbor and repeat the step 2 until no new neighbors are found.
5. Repeat from the step 1 choosing another arbitrary point and increasing the cluster count by 1 until all the points are labelled.

### 4.4.1.1 MinPts determination

Thus, the separation of physical and non-physical modes identified as clusters and outliers respectively, are directly influenced by MinPts. Higher values are usually better for wide range of model orders or noisily dataset where the physical modes cannot be distinguished smoothly. In this dissertation the value of MinPts is selected automatically based on a cluster evaluation metric. Basically, the DBSCAN algorithm is iterated increasing by one the value of MinPts and recording the silhouette index for each iteration. Generally, it's assumed that the number of physically meaningful poles for each physical modes should be greater than 20% of the number of model orders analyzed and similarly the maximum MinPts can't be higher than the dimension of the bigger cluster. In the algorithm the iteration stops when MinPts is so high that DBSCAN can't find any cluster. It can be express mathematically as:

$$20\% \text{ model orders} \leq \text{MinPts} \leq \text{dimension of the bigger cluster} \quad (4-50)$$

Since the DBSCAN has a notion of noise, the outliers are gathered as a further group and can be identified as negative values by the silhouette.

Theoretical basis of the silhouette and an extensive explanation of the procedure are described below.

#### 4.4.1.1.1 Silhouette index

The evaluation of the results obtained by clustering algorithms is a key problem in unsupervised machine learning. Many authors established clustering evaluation indices based on some similarity or dissimilarity measures such as the distance between cluster points for measuring the quality of the division of data set in clusters.

In 1987, Peter J. Rousseeuw [21] proposed the silhouette index based on the comparison of separation and compactness of the clusters calculating the average intra-cluster distance (a) and the average nearest-cluster distance (b) for each point.

This validation method is defined as follow [26]. Assuming that the data set  $D$  is divided in  $K$  clusters denoted as  $\{C_1, C_2, \dots, C_i, C_K\}$ , where  $C_i$  represents the  $i^{th}$  cluster with  $i = \{1, 2, 3, \dots, K\}$ . For the point  $X_i \in C_i$ , let the intra-cluster distance  $a(X_i)$  be the mean of the distances between  $X_i$  and all the other points  $X_j$  belonging to the same cluster  $C_i$ . The lower  $a(X_i)$  is, the lower is the distance between  $X_i$  and the centroid of its cluster and better is its assignation. It is mathematically expressed as:

$$a(X_i) = \frac{1}{N_i-1} \sum_{X_j \in C_i, i \neq j} d(X_i, X_j) \quad (4-51)$$

where  $d(X_i, X_j)$  is the distance between the point  $X_i$  and  $X_j$  in the cluster  $C_i$  and  $N_i$  is the number of points belonging to the cluster  $C_i$ .

The average nearest-cluster distance  $b(X_i)$  is the average distance between the point  $X_i$  belonging to the cluster  $C_i$  and all the points belonging to the neighboring cluster  $C_k$ . It's defined as:

$$b(X_i) = \frac{1}{N_k} \sum_{i \in C_i, j \in C_k} d(X_i, X_j) \quad (4-52)$$

where  $d(X_i, X_j)$  is the distance between the point  $X_i$  and  $X_j$  and  $N_k$  is the number of points in the respective cluster.

Therefore, the silhouette index for the point  $X_i$  is defined as:

$$S(X_i) = \frac{b(X_i) - a(X_i)}{\max(a(X_i), b(X_i))} \quad (4-53)$$

Which can be also written as:

$$S(X_i) = \begin{cases} 1 - \frac{a(X_i)}{b(X_i)}, & \text{if } a(X_i) < b(X_i) \\ 0, & \text{if } a(X_i) = b(X_i) \\ \frac{b(X_i)}{a(X_i)} - 1, & \text{if } a(X_i) > b(X_i) \end{cases} \quad (4-54)$$

Consequently, for each point  $X_i$  can be assumed:

$$-1 \leq S(X_i) \leq 1 \quad (4-55)$$

Worthy of note is the case with dataset divided in only one cluster, in which case  $S(X_i) = 0$ .

As  $a(X_i)$  is representative of similarity and  $b(X_i)$  of dissimilarity, a well-matching of  $X_i$  to its cluster requires  $a(X_i) \ll b(X_i)$  and value of  $S(X_i)$  close to 1. Contrarily, low value of  $b(X_i)$  means that  $X_i$  would be better clustered if it was assigned to its neighboring cluster. If  $S(X_i)$  is close to zero means that  $X_i$  lies in the border region between its cluster and the closest one.

The silhouette index is also calculated for clusters for measuring how well grouped are the points in that cluster. The silhouette index for clusters is defined as:

$$S(C_i) = \frac{1}{N_i} \sum_{X_i \in C_i} S(X_i) \quad (4-56)$$

where  $N_i$  is the number of points in the cluster  $C_i$  and  $S(X_i)$  is the silhouette index calculated for each point  $X_i$  belonging to the cluster  $C_i$ .

Furthermore, the silhouette index calculated for all the data set measures how well the data have been clustered. Besides, the application in machine learning data analysis is not only limited on the evaluation of clusters per se, but it can be used for selecting parameters on the clustering algorithm. For example, for choosing the number of clusters  $k$  on K-means clustering or, as in our study, for selecting one parameter of the DBSCAN clustering algorithm.

In this work, the silhouette coefficient corresponds to a certain MinPts and is calculated for each iteration of DBSCAN. The MinPts correlated to the maximum

silhouette coefficient is chosen as the best fitting. In this way, this parameter is set in base of the performance of the evaluation metric and the user doesn't need to specify it manually. Referring to the other works that use the hierarchical clustering instead of the DBSCAN, the silhouette coefficient varies from 0.6 to 1 in most of the studied cases for the presence of outliers.

### 4.4.1.2 Epsilon determination

The DBSCAN clustering also requires the neighborhood clustering threshold epsilon (eps). In our algorithm eps is computed by means of the k-nearest neighbor distance (k-NN distance) and of a heuristic method called the elbow rule [28]. Finding the most suitable epsilon for a certain data set is a key feature that directly influence the results of the clustering. Smaller values of eps would split the physical modes into several modes or even dismiss them, whereas larger values would produce broader clusters constituted also by outliers or even by several physical modes.

As a result, the explained below elbow rule provides a good way to automatize the DBSCAN clustering algorithm that is able to handle very different data set as reported in the next chapter.

#### 4.4.1.2.1 Elbow rule

The elbow rule is computed as follows.

For each pole  $X_i$ , let  $D_i$  be the distance between  $X_i$  and its  $k^{th}$  nearest neighbor pole. The vector  $D = \{D_1, D_2, \dots, D_i, D_N\}$ , constituted by  $N$  k-nn distances, where  $N$  corresponds to the number of points in the data set, is then plotted in ascending order on a k-distance graph. As expressed previously, outliers are poles that lie in low-density region, thus their  $D_i$  are higher with respect the ones of the physical poles. The aim of the elbow rule is to find the knee of the curve that precedes the sharp increase of k-nn distances due to the presence of the outliers. This point of maximum curvature would correspond to the best epsilon. In this way the DBSCAN should discard the most isolated poles during the clustering. The knee of the curve,

or the point of maximum curvature, is evaluated as the farthest point to the line that connects the first and the last point of the curve. Given that this method is heuristic, it is not applicable in data sets with scattered poles with no clearly defined clusters where the k-nn distance would not be represented by a clear curve. Fortunately it's not the case of the stabilization diagrams obtained in the automated operation modal analysis. Since the physical modes are represented by compacted and sometimes not far clusters, as  $k$  for computing the k-nn distance is chosen the second-nearest neighbor distance. In this way, the physical poles are clearly separated even with similar parameters and high-noise conditions. However, further details and examples are reported in the following chapters.

#### 4.4.2 Distance vector

The DBSCAN and the k-nn distances can be performed with any distance function. The definition of similarity, thus the distance between poles, can therefore be seen as an additional parameter. Worthy of remark, the physical poles are distinguished from the computational ones for their similarity at different model orders. The min-max normalized difference of eigenfrequencies (eq. 4-31) and the modal assurance criterion (MAC) (4-36) are used as comparison parameters. Differently to what computed by Reynders and all [2], the difference of eigenfrequencies is preferred than the difference of eigenvalues because damping estimates are jeopardized by large error bounds, related probably to the non-linear behavior of damping. The distance matrix that characterizes all the poles of all the model orders is defined as follows.

Let  $D_{i,j}$  be the distance between the pole  $i$  and the pole  $j$  computed as:

$$D_{i,j} = \frac{\Delta f_{i,j} - \min(\Delta f_{i,j})}{\max(\Delta f_{i,j}) - \min(\Delta f_{i,j})} + \frac{v_{i,j} - \min(v_{i,j})}{\max(v_{i,j}) - \min(v_{i,j})} \quad (4-57)$$

Where:

$$\Delta f_{i,j} = |f_j - f_i| \quad v_{i,j} = 1 - MAC(\phi_i, \phi_j) \quad (4-58)(4-59)$$

Clearly, the matrix  $D_{i,j}$  is symmetric with null values on the diagonal that is mathematically expressed as:

$$D_{i,j} = D_{j,i}, \quad D_{i,j} = 0 \text{ if } i = j \quad (4-60)(4-61)$$

Consequently, the dimension of the distance matrix  $D_{i,j}$  is  $L * L$ .

In order to find the  $k^{th}$  value used in the elbow rule, the matrix D is calculated, and each row is sorted in ascending order. The  $k^{th}$  value corresponds to the k-nn neighbor of the pole  $i$ . Differently, for computing the DBSCAN the number of terms of the distance matrix is  $L * \frac{(L-1)}{2}$  because only the lower triangular part of the matrix is considered to avoid useless duplicates.

#### 4.5 Final modal parameter selection

The DBSCAN algorithm already identifies the physical modes as distinct clusters separating the physical poles from the outliers. Each cluster contains a large number of poles characterized by their own natural frequencies, damping ratios and eigenvectors. Hence, the last part of the algorithm revolves around the choice of which modal parameters better represent the physical clusters. Several strategies that rely on different statistical tools can be retrieved from literature, in this work five alternatives are tested and compared during numerical and experimental case studies. However, this selection doesn't substantially influence the final results. For each physical cluster the modal parameters are estimated with the following methods:

1. Magalhães et al. [22] suggests averaging natural frequencies, damping ratios and mode shapes calculated from all observations in each physical cluster. The modal parameters obtained are finally considered as physical modes.
2. Reynders et al. [2] chose the mode of the physical pole that has closest damping ratio  $\xi$  to the average damping ratio  $\bar{\xi}$  of the poles belonging to the same cluster.

3. The modal parameters of the pole that has the minimum cumulative distance between all its modal parameters and the mean of all the parameters of the poles belonging to the same cluster are chosen.
4. Choosing the modal parameters of the pole that has the most similar eigenvectors to the mean of the eigenvectors of all the poles belonging to the same cluster. The similarity is evaluated as sum of the differences between eigenvectors for each DoF.
5. Schwochow and Jelcic [23] suggests using the modes from the lowest possible model order, which still has an observation in each physical cluster.

The above-mentioned methods present advantages and disadvantages which depends on the application of the algorithm. All of them can be used indifferently in the presented AOMA algorithm. However, results are showed in following chapters.

### **4.6 Major differences with previous proposals**

In recent years, in the field of OMA, many researchers focused on the development of automated, robust, and efficient methods for the elimination of spurious modes and the recognition of closely spaced modes previously identified by System Identification (SI) techniques. Generally, this challenge is usually addressed following three stages that corresponds to the manual identification: clearing out the stabilization diagram, detecting columns of stable poles and determining the final modal parameters of each column. Although different approaches have been implemented, most of them rely on the hierarchical clustering algorithm for detecting columns of stable poles. In the presented AOMA method a novel approach based on an automated DBSCAN is proposed.

Actually, DBSCAN has been already implemented in the field of AOMA by C. Ye and X. Zhao [24] in 2020. Nevertheless, the selection of MinPts and Epsilon has been made according to the range of model orders, thus the performance of the algorithm depended on the dimension of the data set. In the proposed AOMA, the selection of these parameters relies on a cluster validation criteria and on a heuristic method that

evaluate the quality of the clusters as reported in **DBSCAN**. Therefore, the clustering obtained iterating all the MinPts included in a range are evaluated and the best performing is chosen as the final one. This approach allows to correctly identify the modal parameters independently to the type and dimension of the data set .

The second characteristic that distinguishes the presented AOMA to most of the AOMA algorithms is the use of absolute differences instead of relatives differences. This choice has been already made by Mugnaini [4] due to the well-known issue that the relative differences can lead to an underestimation of the entities represented by low values.

Differently to the other AOMA methods the block rows of the Henkel matrix has been set according to Reynders and De Roeck's suggestions [18] as follow:

$$i = \frac{f_s}{2 f_0}$$

The rest of the algorithm is derived from other AOMA techniques reported in **References** except for small differences explained in **Methodology**.

## 5 Numerical case

The proposed AOMA method is firstly tested on a numerical case with the aim to evaluate the capability to identify the correct modal parameters. The numerical simulation is representative of a three-storeys multi-bay shear type frame and its results are compared with the modal parameters calculated with the eigenproblem attesting the robustness of the AOMA algorithm. Besides, the signal length as well as the level of noise artificially added to the signal are variated to evaluate their influence on the results.

### 5.1 Description

The three-storeys multi-bay shear type frame presents the following characteristics:

- The frame has 3 degrees of freedom for each floor: two orthogonal displacements and one rotation. As results, the frame has 9 degrees of freedom in total.
- The frame is considered as shear type, thus each floor is considered as infinitely rigid, the mass is concentrated on the floors and the columns are flexible and stiff.
- Since the frame is not symmetric but it's irregular, there is an eccentricity between the center of resistance and the center of mass.
- In accordance with the system shown in *Figure 3-1* the mass matrix is the following:

$$\begin{bmatrix} 199926 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & -1262530 & 0.0000 & 0.0000 \\ 0.0000 & 173425 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & -1095179 & 0.0000 \\ 0.0000 & 0.0000 & 200717 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & -1267529 \\ 0.0000 & 0.0000 & 0.0000 & 199926 & 0.0000 & 0.0000 & 7.4000 & -3.1100 & 0.0293 & \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 173425 & 0.0000 & 0.0000 & 659897 & 0.0000 & \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 200717 & 0.0000 & 0.0000 & 772463 & \\ -1262530 & 0.0000 & 0.0000 & 762350 & 0.0000 & 0.0000 & 199926 & 0.0000 & 0.0000 & \\ 0.0000 & -1095179 & 0.0000 & 0.0000 & 659897 & 0.0000 & 0.0000 & 173425 & 0.0000 & \\ 0.0000 & 0.0000 & -1267529 & 0.0000 & 0.0000 & 772463 & 0.0000 & 0.0000 & 200717 & \end{bmatrix} [Kg]$$

- In accordance to the following matrix:

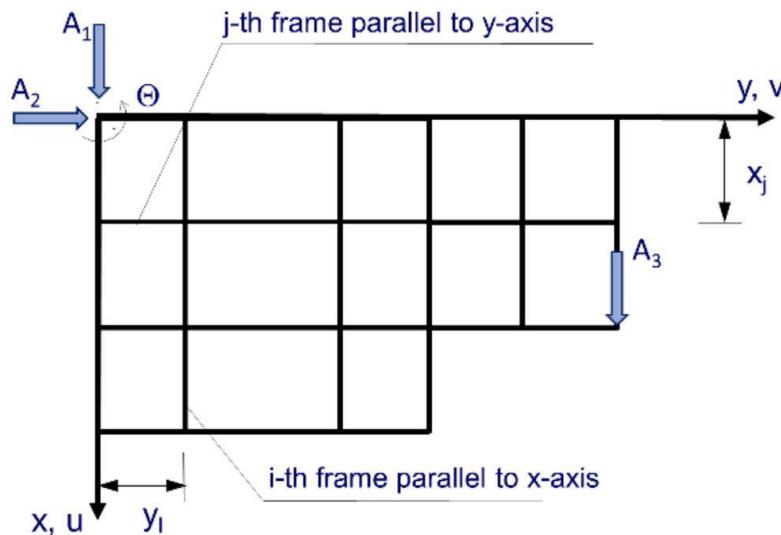
## Numerical case

$$\begin{bmatrix} k_{xx} & k_{xy} & k_{x\theta} \\ k_{yx} & k_{yy} & k_{y\theta} \\ k_{\theta x} & k_{\theta y} & k_{\theta\theta} \end{bmatrix}$$

The stiffness matrix is the following:

$$\begin{bmatrix} 1.8900 & -0.8630 & -0.0576 & 0.0000 & 0.0000 & 0.0000 & -13.4000 & 6.0900 & -0.4060 \\ -0.8630 & 1.3600 & -0.5740 & 0.0000 & 0.0000 & 0.0000 & 6.0900 & -9.5600 & 4.0500 \\ -0.0576 & -0.5740 & -0.4410 & 0.0000 & 0.0000 & 0.0000 & -0.4060 & 4.0500 & -3.1100 \\ 0.0000 & 0.0000 & 0.0000 & 1.9300 & -0.7780 & 0.0022 & 7.4000 & -3.1100 & 0.0293 \\ 0.0000 & 0.0000 & 0.0000 & -0.7780 & 1.3300 & -0.5200 & -3.1100 & 5.2500 & -2.0400 \\ 0.0000 & 0.0000 & 0.0000 & 0.0022 & -0.5200 & 0.5110 & 0.0293 & -2.0400 & 2.0300 \\ -13.4000 & 6.0900 & -0.4060 & 7.4000 & -3.1100 & 0.0293 & 21.1000 & -9.6400 & 0.4700 \\ 6.0900 & -9.5600 & 4.0500 & -3.1100 & 5.2500 & -2.0400 & -9.6400 & 15.5000 & -6.3700 \\ -0.4060 & 4.0500 & -3.1100 & 0.0293 & -2.0400 & 2.0300 & 0.4700 & -6.3700 & 5.4000 \end{bmatrix} 10^9 [Nm^2]$$

- The sensors are placed following the picture reported below.



(Same plan view for all stories)

*Figure 5-1 Disposition of accelerometers*

- The damping ratios associated to each mode are the following:

$$dr_1 = 0.20\%$$

$$dr_2 = 0.60\%$$

$$dr_3 = 1.30\%$$

$$dr_4 = 3.50\%$$

$$dr_5 = 0.80\%$$

$$dr_6 = 1.50\%$$

$$dr_7 = 0.70\%$$

$$dr_8 = 2.90\%$$

$$dr_9 = 2.00\%$$

The system is excited with an input signal generated by random sampling with a Gaussian distribution  $\mathcal{N}(0,0.005)$  which is implemented with the Matlab command “*randn*”. In this way the signal can be considered as white noise. The time is discrete, i.e. the time vector and the signal are defined for each time interval  $1/f_s$ .

The time vector  $t$  used to define the signal is defined as:

$$t = \begin{pmatrix} 0 \\ 1/f_s \\ 2/f_s \\ \dots \\ n/f_s \end{pmatrix}$$

where  $f_s$  is the sampling frequency and  $n+1$  are the components of the vector. Thus, the signal has a duration equal to  $n/f_s$ . In the reported case study, the signal has a duration of 180 sec and a sampling frequency of 100 Hz.

Therefore, the time vector  $t$  becomes:

$$t = \begin{pmatrix} 0 \\ 0.01 \\ 0.02 \\ \dots \\ 180 \end{pmatrix}$$

In order to obtain an acceleration similar to the real one, the signal has been scaled with a coefficient equal to 0.005. Then, the Gaussian distribution acquired the following variance:

$$\sigma = 0.005$$

The signal is applied to the origin O and the excitation is transmitted along all the degree of freedom with different intensities through the use of the following track vector:

$$track = \begin{pmatrix} 0.6761 \\ 0.1690 \\ 0.1690 \\ -0.5070 \\ 0.3380 \\ 0.3380 \\ -0.0101 \\ -0.0101 \\ 0.0034 \end{pmatrix}$$

Since the geometry of the system and the excitation are calculated, the displacements are computed for each floor from O with the following integral:

$$u_i = - \left( \frac{\psi_i}{\sqrt{\psi_i' M \psi_i}} \right)' M T_r \frac{1}{\omega_i \sqrt{1 - (dr_i)^2}} \int_0^T wn_i(t) * (e^{-dr_i \omega_i t} * \sin(\omega_i \sqrt{1 - (dr_i)^2} t)) dt$$

where

- $M$  is the mass
- $\psi_i, \omega_i, dr_i$  are the eigenvector, the angular frequency and the damping ration of the  $i^{th}$  mode
- $T_r$  is the track vector
- $wn_i(t)$  is the acceleration applied to the system
- $t$  is the time vector
- $T$  is the duration of the applied acceleration

Since the time is discrete the integration is considered as a summation.

As result the output signal is defined for each accelerometer.

The artificial addition of noise is due to assess the capability of the AOMA algorithm to identify the modal parameters of the system in more similar to the reality conditions. The noise is added to the signal:

$$disturbed\ signal = signal + noise\ perc * (Std\ Dev(signal))$$

## 5.2 Automated Operational Modal Analysis

The described numerical simulation is analysed with the AOMA algorithm.

The signal setup is the following:

- Sampling frequency: 100 Hz
- Signal duration 300 s
- Noise 5 %

The SSI-COV is set as follows:

- Range of model orders 10-80
- Block rows of the Henkel matrix  $f_s/2$

The range of model orders is chosen on the basis of studies conducted by Mugnaini et al [4]: the minimum model order is set to  $n_{min} = 10$  because one order higher than the degree of freedoms of the system and the maximum model order is equal to 80 because it shows higher values of Precision and Recall.

### 5.2.1 Hard validation criteria

The first cleanliness of the stabilization diagram obtained from the SSI method relies on the meaning of the modal parameters. Only the poles with the following characteristics have been considered as possible physical:

$$0\% \leq \xi_j \leq 20\%$$

$$Re(\lambda_j) \geq 0, \quad Im(\lambda_j) \neq 0$$

Before applying the HVC for each model order the number of poles identified is equal to half the dimension of the model order. Considering that the model order is increased by two and a range of model order between 10 and 80 is set, the poles identified by the SSI are 810.

## Numerical case

---

After applying the HVC the residual poles are 606.

The stabilization diagrams obtained before and after the application of the HVC are reported below.

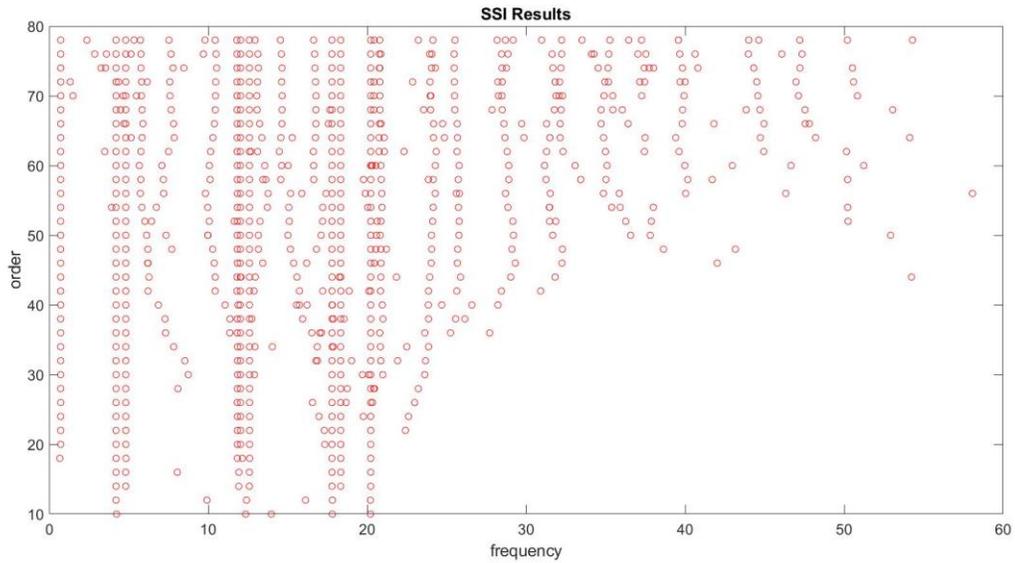


Figure 5-2 Stabilization diagram obtained before HVC

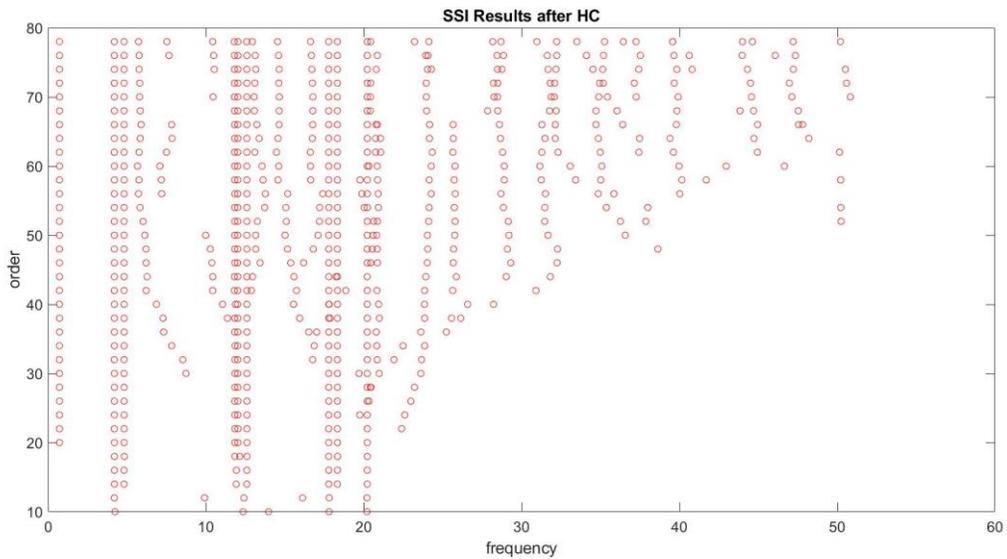


Figure 5-3 Stabilization diagram obtained after HVC

### 5.2.2 Soft Validation Criteria

Referring to the methodology described in **Methodology**, the distinction between stable and unstable poles is addressed by means of Soft Validation Criteria.

The results of the k-means clustering based on the feature vector are shown in the lower left triangular of the *Table 5-1*. The upper right triangular represents the correlation coefficients between individual features: higher is the value, greater is the influence on the clustering process.

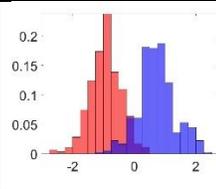
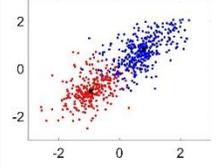
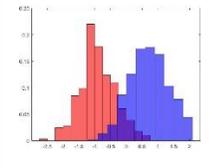
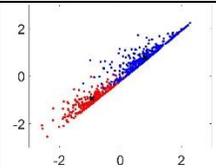
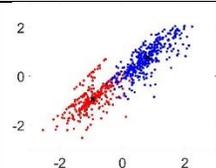
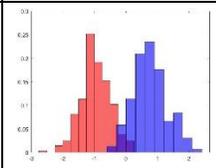
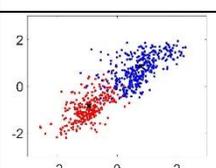
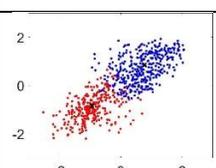
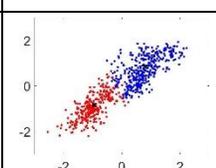
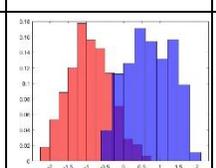
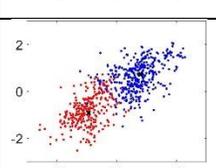
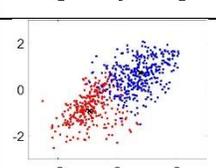
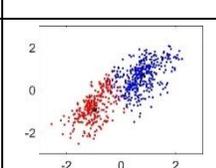
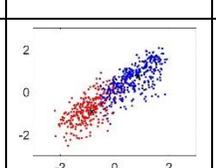
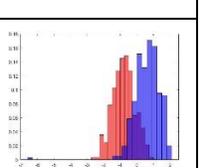
$\Delta f$		0.99	0.87	0.95	-0.12
$\Delta \xi$			0.97	0.98	-0.3
$\Delta \lambda$				0.99	-0.05
$\nu$					-0.43
$\Delta MPD$					
	$\Delta f$	$\Delta \xi$	$\Delta \lambda$	$\nu$	$\Delta MPD$

Table 5-1 Influence of each parameters on clustering process

## Numerical case

The stabilization diagram and damping ratio-model order diagram of the stable poles identified after the SVC are reported as well. The red circled poles are recognized as stable, whereas the black cross poles as unstable.

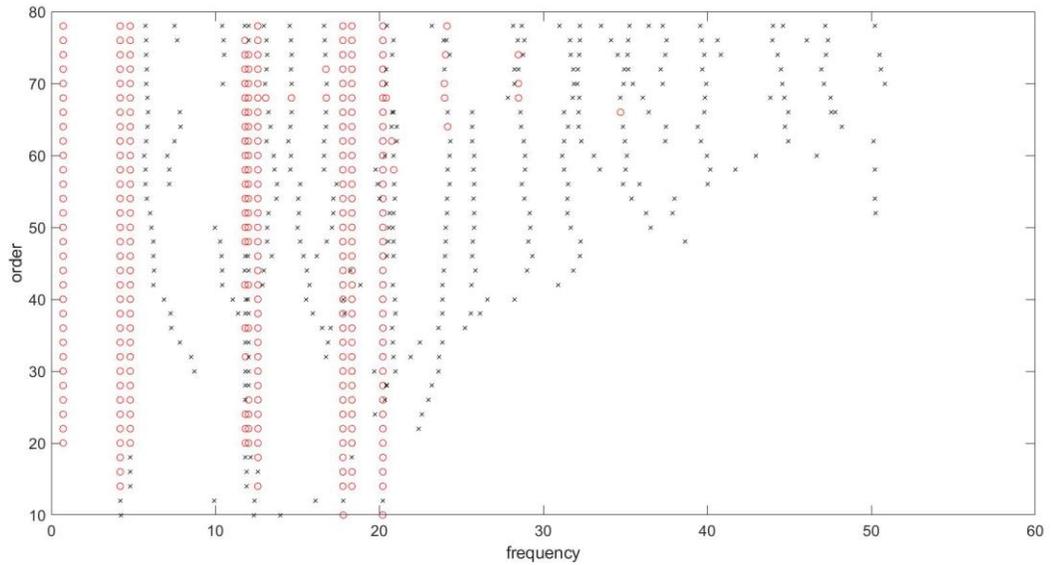


Figure 5-4 Stabilization diagram after SVC

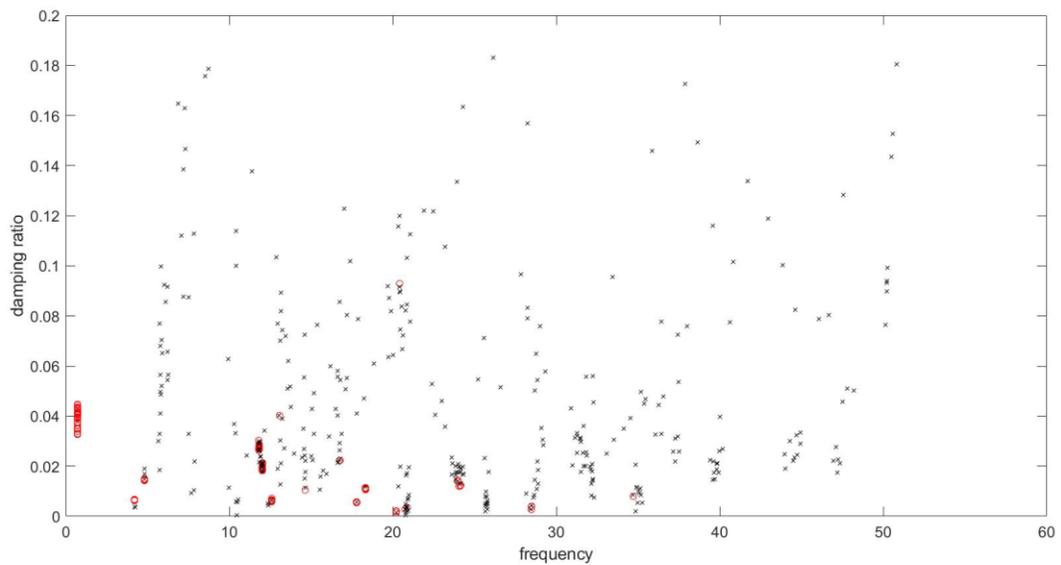
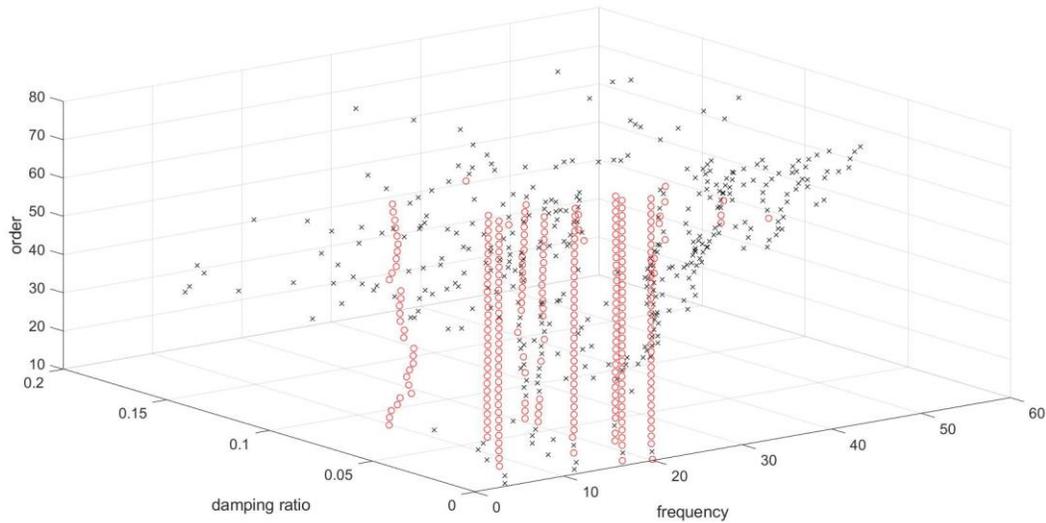


Figure 5-5 Damping ratio-model order diagram after SVC

The 3-D diagram that shows the combination of model order, damping ratio and frequency is reported in *Figure 5-6*.



*Figure 5-6 Damping ratio-model order-frequency diagram after SVC*

The red circled (stable) poles which are successively processed by the DBSCAN are 284.

### 5.2.3 DBSCAN

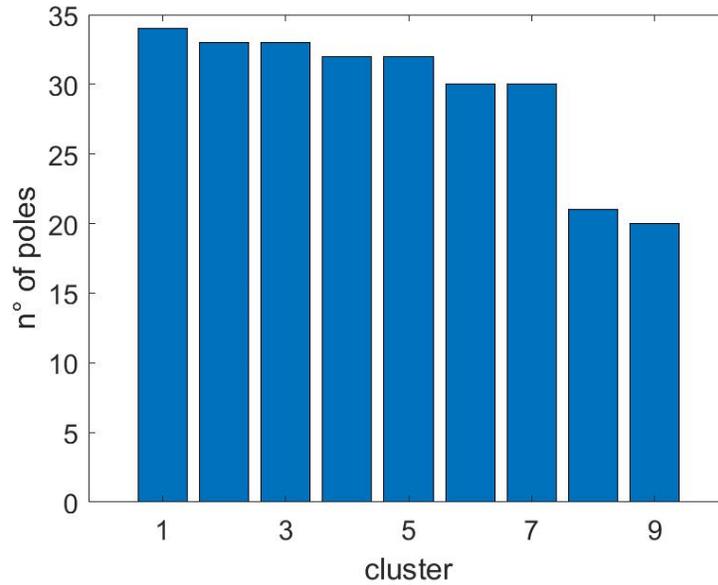
The stable poles are grouped together in base of their similarity in order to obtain the clusters representative of the modal parameters of the structure.

As already explained, the DBSCAN clustering algorithm relies on the definition of two parameters that are automatically calculated.

The 284 stable poles are merged in 9 distinct clusters which contain at least 20 poles each. In *Figure 5-7* the dimension of each group is shown.

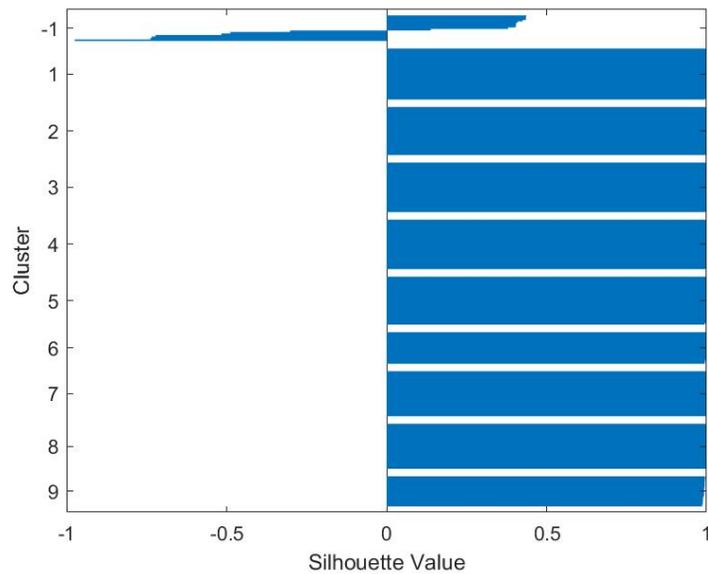
## Numerical case

---



*Figure 5-7 Number of poles in each cluster*

As shown in Figure 5-8, the silhouette index assumes positive values which are very close to 1 for all the poles merged in clusters. Being the outliers scattered, they are grouped in the cluster -1 that is represented by varying values of silhouette.



*Figure 5-8 Silhouette values*

## Numerical case

Finally, the stabilization and the order-frequency-damping ratio diagrams are reported. Each cluster is represented by different colours, whereas the outliers are cross points.

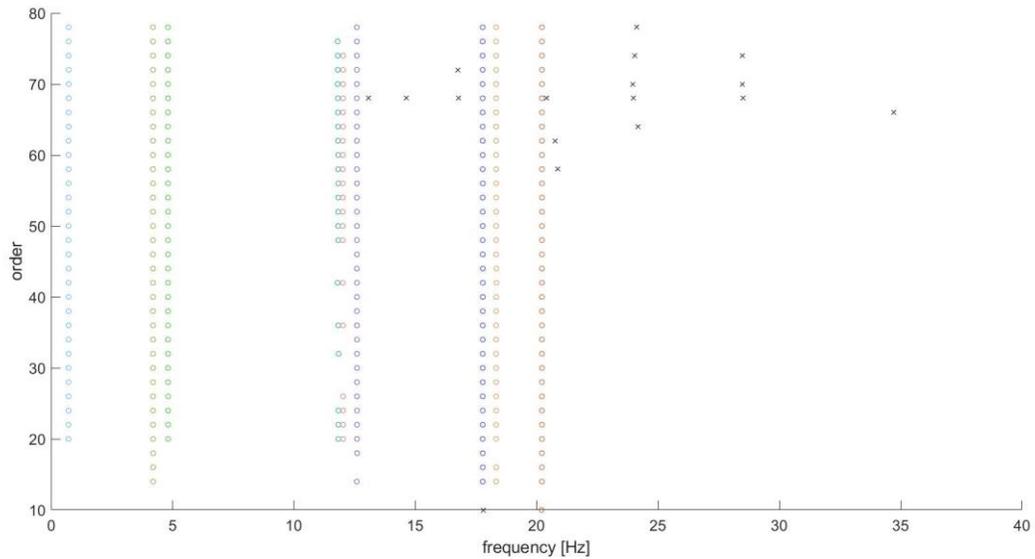


Figure 5-9 Stabilization diagram after DBSCAN

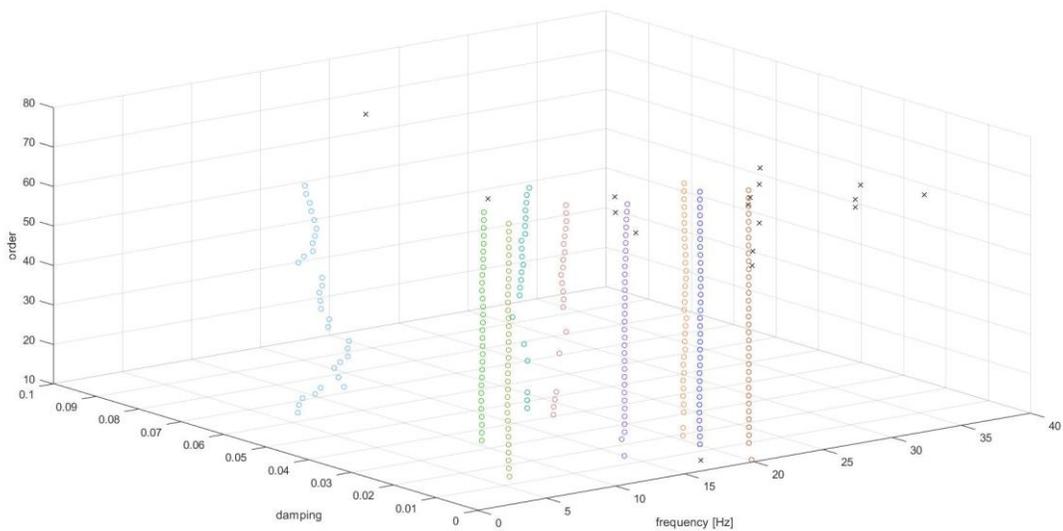


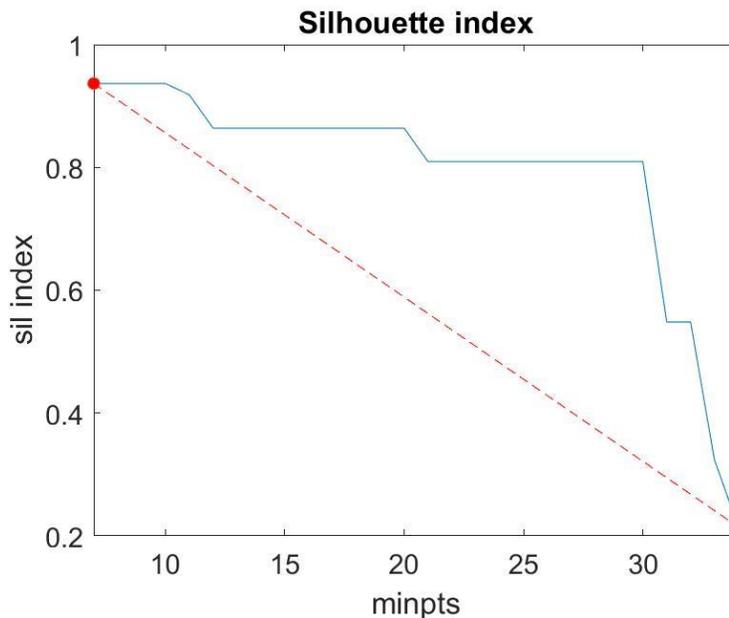
Figure 5-10 Damping ratio-model order-frequency diagram after DBSCAN

### 5.2.3.1 MinPts determination

The silhouette index is an evaluation metric of validation of consistency within clusters of data. The DBSCAN is iterated increasing by one the value of MinPts and calculating the silhouette index for each iteration.

Generally, higher is Minpts lower is the number of identified clusters and consequently higher is the number of outliers. Therefore, the silhouette index usually decreases with higher MinPts.

In *Figure 5-11* the best silhouette index corresponds to the lower value of MinPts investigated, i.e. MinPts = 4. The evaluated MinPts range between 4 to 35 following the equation (4-50).

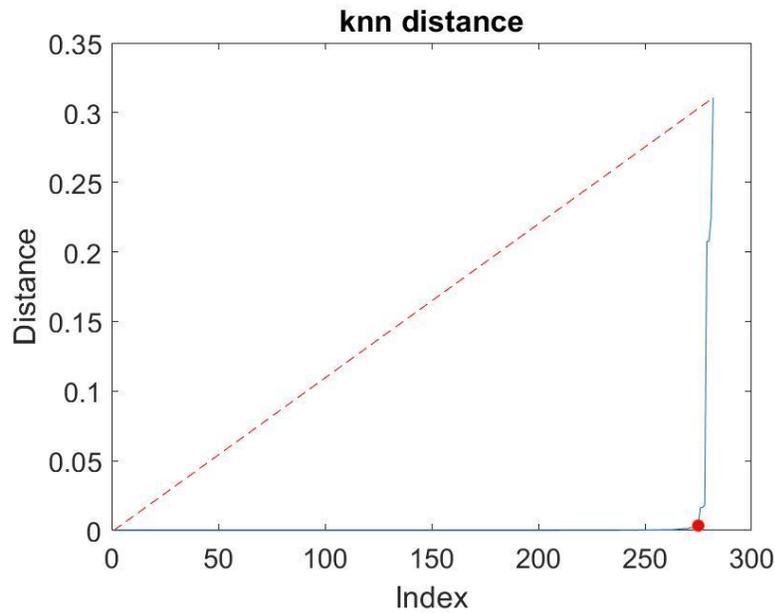


*Figure 5-11 MinPts determination*

### 5.2.3.2 Epsilon determination

The epsilon parameter is calculated using the k-nearest neighbour distance and the heuristic elbow rule method. The second neighbour distance of all the 284 poles is calculated and sorted in ascending order. The point of maximum curvature

corresponding to the red dot in *Figure 5-12* is considered as epsilon. In this case study it is equal to 0.0035.



*Figure 5-12 Epsilon determination*

### 5.2.4 Final modal parameters

For each cluster the final modal parameters are calculated using the five different methods explained in *Final modal parameter selection*. The results are summarized in the following table.

	Type of mode	Method 1 Identification [Hz]	Method 2 Identification [Hz]	Method 3 Identification [Hz]	Method 4 Identification [Hz]	Method 5 Identification [Hz]	FEM Identification [Hz]
$f_1$	1th Bending along X	0.7245	0.7227	0.7236	0.7234	0.7229	0.7277
$f_2$	1th Torsional	4.2039	4.2038	4.2037	4.2037	4.2040	4.1979

## Numerical case

---

$f_3$	1th Bending along Y	4.8200	4.8195	4.8195	4.8211	4.8221	4.8239
$f_4$	2th Bending along X	11.8248	11.8270	11.8227	11.8227	11.8379	11.8675
$f_5$	2th Bending along Y	12.0262	12.0195	12.0286	12.0286	12.0329	12.0137
$f_6$	3th Bending along Y	12.6012	12.6008	12.6007	12.6007	12.6039	12.5996
$f_7$	2th Torsional	17.7858	17.7860	17.7848	17.7848	17.7859	17.7983
$f_8$	4th Bending along Y	18.3361	18.3335	18.3366	18.3455	18.3353	18.3285
$f_9$	3th Torsional	20.2238	20.2235	20.2237	20.2237	20.2238	20.2294

*Table 5-2 identification results of numerical case*

The choice of the method used to define the final modal parameters doesn't not affect considerably the results. The maximum difference among the frequencies is indeed about 1%. Besides, all the modes identified by the FEM are recognized by the proposed AOMA as well.

In *Figure 5-13* to *Figure 5-21* the mode shapes identified by the AOMA and by the eigenproblem are compared. The continuous lines refer to the AOMA identification, whereas the dashed lines to the eigenproblem identification.

# Numerical case

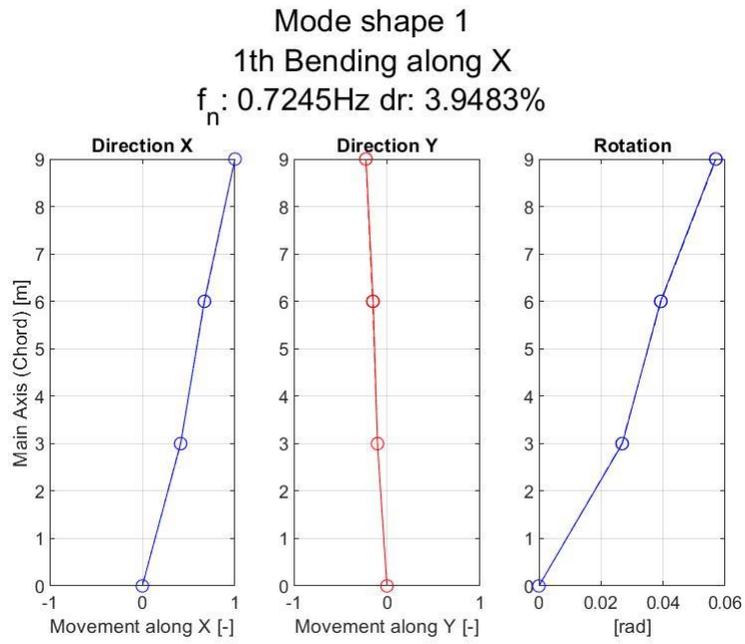


Figure 5-13 1<sup>st</sup> mode shape identified

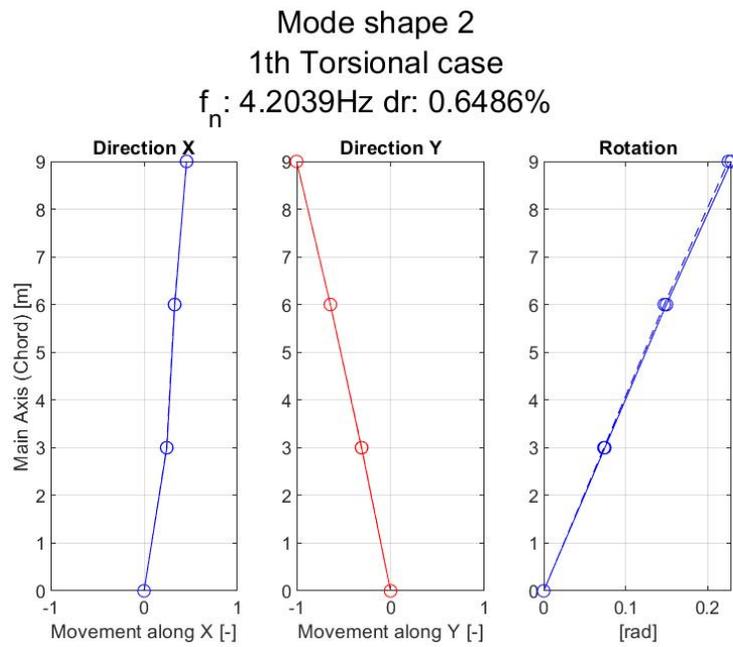


Figure 5-14 2<sup>nd</sup> mode shape identified

Numerical case

Mode shape 3  
1th Bending along Y  
 $f_n$ : 4.8200Hz dr: 1.4583%

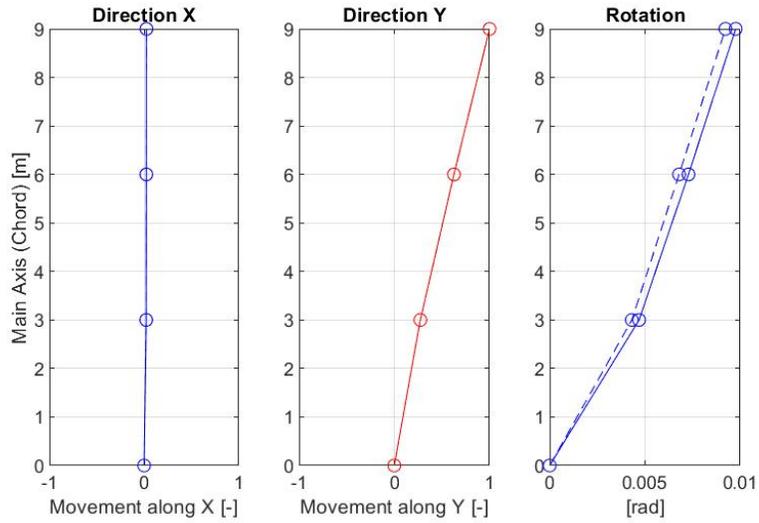


Figure 5-15 3<sup>rd</sup> mode shape identified

Mode shape 4  
2th Bending along X  
 $f_n$ : 11.8248Hz dr: 2.7675%

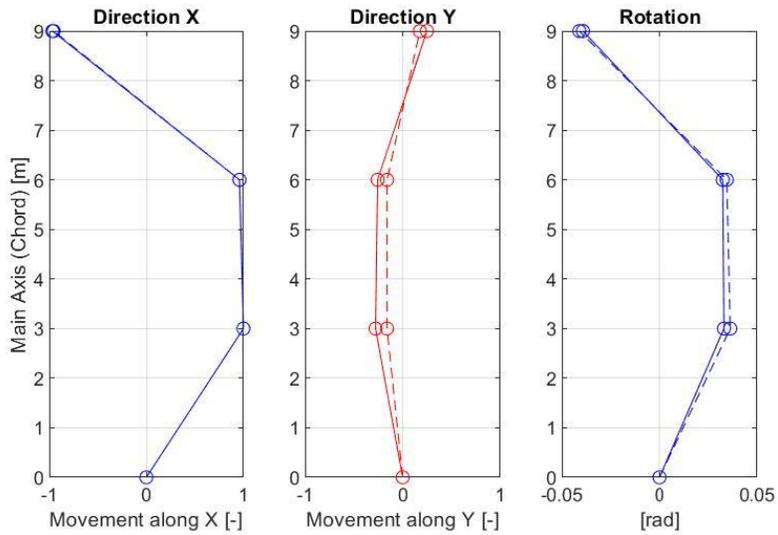


Figure 5-16 4<sup>th</sup> mode shape identified

Numerical case

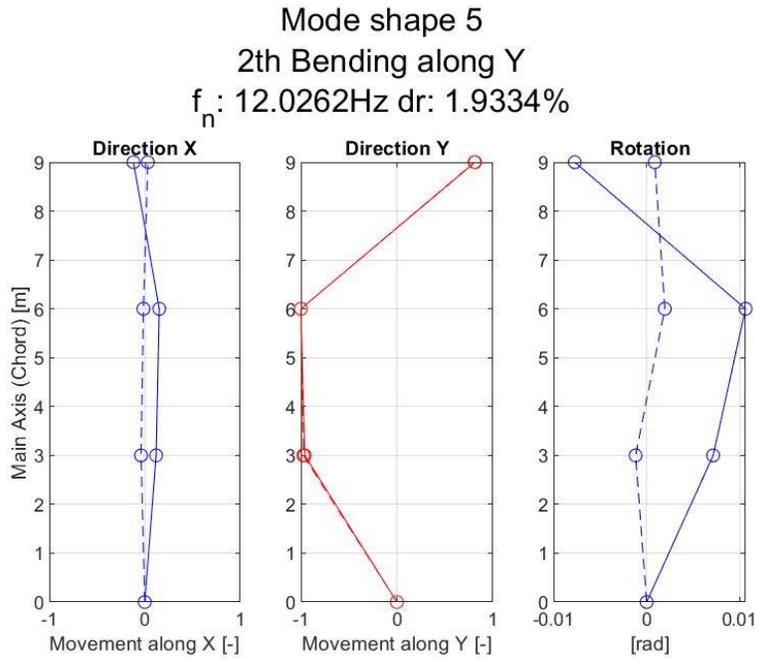


Figure 5-17 5<sup>th</sup> mode shape identified

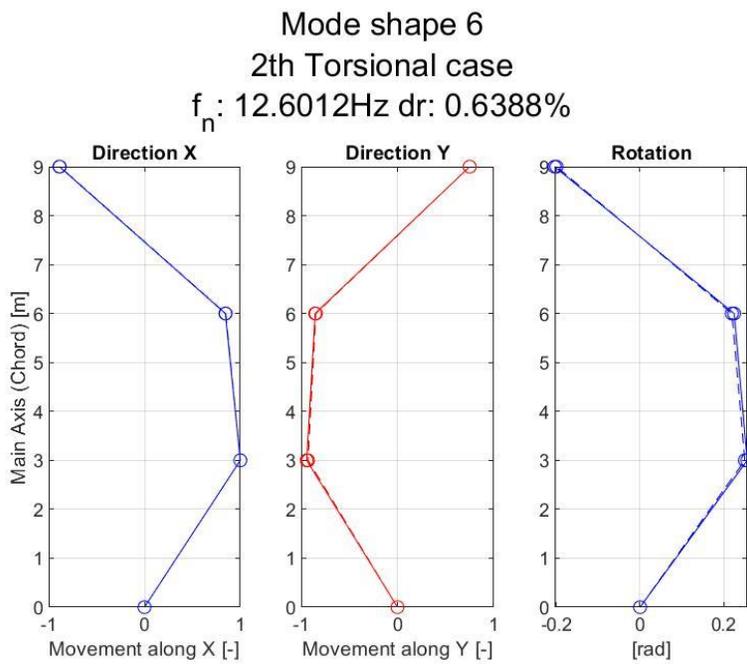


Figure 5-18 6<sup>th</sup> mode shape identified

Numerical case

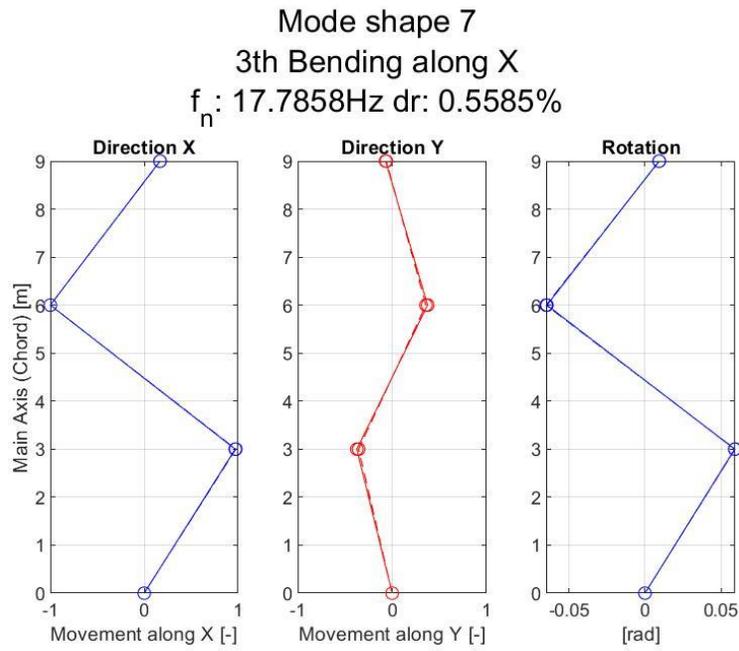


Figure 5-19 7<sup>th</sup> mode shape identified

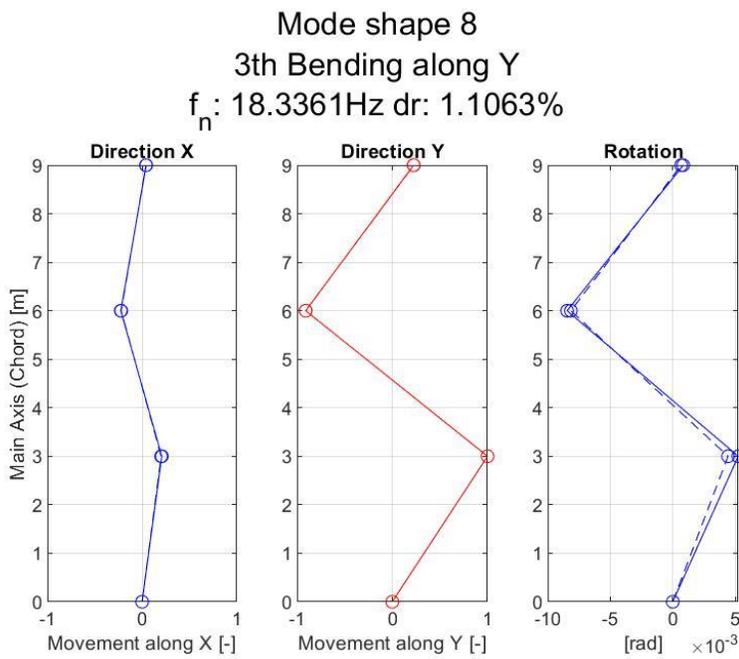


Figure 5-20 8<sup>th</sup> mode shape identified

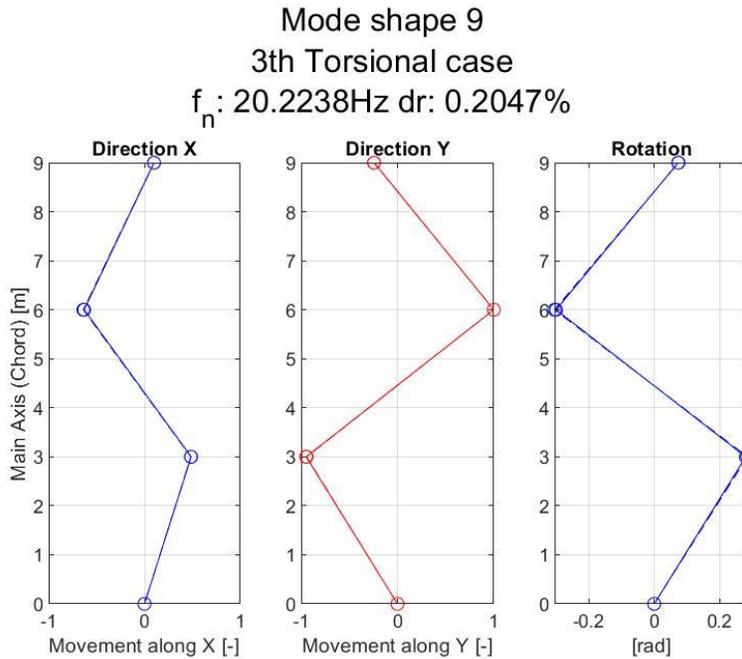


Figure 5-21 9<sup>th</sup> mode shape identified

Basically, the mode shapes identified by the eigenproblem and by the AOMA method result very similar except for the rotation of mode shape 5. Actually, the difference between the rotations is about 0.008 rad, thus it can be considered neglectable.

The type of the mode shape, i.e. bending or torsional case, is recognized automatically relying on the following assessments:

<b>Condition</b>	<b>Mode shape type</b>
$0.4 \text{ mean}(\text{abs}(u)) > \text{mean}(\text{abs}(v))$	Bending along x
$0.4 \text{ mean}(\text{abs}(v)) > \text{mean}(\text{abs}(u))$	Bending along y
Otherwise	Torsional

Table 5-3 Conditions to determinate the type of mode shape

where  $u$  and  $v$  are the mode shape components along x and y.

### **5.3 Influence of signal duration and noise**

In this paragraph the signal duration and the noise are varied with the aim of better understand their influence on the identification of final parameters performed by the AOMA method.

The signal durations adopted in this study are: 5, 15, 25, 35, 45, 55, 65, 75, 85, 95, 105 min, whereas the noises added to signal are: 1%, 20%, 40%, 60%, 80%, 100%, 120%, 140%, 160%, 180%, 200%.

In order to quantify the performance of the AOMA three parameters are implemented: the Precision, the Recall and the F1 score.

#### **5.3.1 Precision, Recall and F-score**

In information retrieval, Precision, Recall and F-score are empirical evaluation metrics used to estimate the goodness of a system in retrieving the information requested by the user.

Processing signals with different noise and duration with the AOMA, the modes identified may or may not be representative of the modes of the numerical case, i.e. the modes identified with the eigenproblem. Moreover, the identification can lead to a number of modes different to the number of modes which represent the structure. The Precision, the Recall and the F-score are metrics that evaluate the goodness of the identification. These metrics are based on the definition of these two parameters:

- True positive (TP): The modes identified with the AOMA that corresponds to the real representative modes of the system
- False positive (FP): The modes identified with the AOMA that don't match any real mode of the system

The similarity of the modes is calculated evaluating their frequency and their mode shape independently.

An identified mode is considered true positive in terms of frequency if the difference of its frequency to the frequency of any real representative mode is less than 0.1 Hz. Contrarily, an identified mode is considered as true positive in terms of mode shape if the MAC calculated between the mode and any representative mode of the system is higher than 0.9. Clearly, if these conditions are not satisfied the identified mode is considered as false positive. The above-mentioned criteria are summarized in the following table:

	True Positive	False Positive
Frequency	$ f_r - f_i  \leq 0.1 \text{ Hz}$	$ f_r - f_i  \geq 0.1 \text{ Hz}$
Mode shape	$MAC(\psi_r, \psi_i) \geq 0.9$	$MAC(\psi_r, \psi_i) \geq 0.9$

Table 5-4 Thresholds for acceptance as True Positive (TP)

where  $f_r$  and  $\psi_r$  are the frequency and the mode shape of a real mode that represent the structure, on the other hand  $f_i$  and  $\psi_i$  are the frequency and the mode shape of a mode identified with the AOMA.

Precision, Recall and F-score are defined as follows:

- Precision: ratio between the true positive modes and the total number of modes identified by the AOMA
- Recall: ratio between the true positive modes and the total number of modes representative of the system (in this case study equal to 9)
- F-score: weighted average of Precision and Recall. Thus, this score measures the accuracy of the identification

These scores are calculated as follows:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{9}$$

$$F \text{ score} = 2 * \frac{precision * recall}{precision + recall}$$

All the scores range between 1 which means total similarity between the identified modes and the modes which represent the structure to 0 that corresponds to complete dissimilarity.

### 5.3.2 Results

Considering the noise and the signal duration variations, the total number of cases analysed is equal to 121. Thus, the evaluation scores are calculated 121 times building a grid with dimension 11x11.

The charts are represented in Figure 5-22 to *Figure 5-27* and have the following features:

- The x axis represents the signal duration (ranging from 5 to 105 min)
- The y axis represents the noise variation (ranging from 1 to 200%)
- The value of the score metrics is represented with colours described in the colour legend on the right of the figure
- The score metrics that don't correspond to a node are interpolated

## Numerical case

---

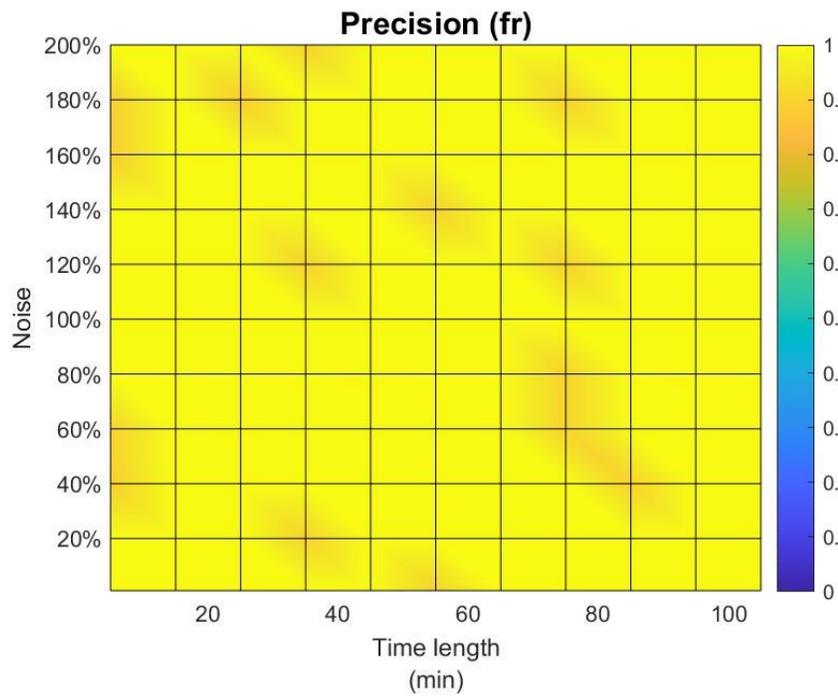


Figure 5-22 Precision in terms of difference of frequencies

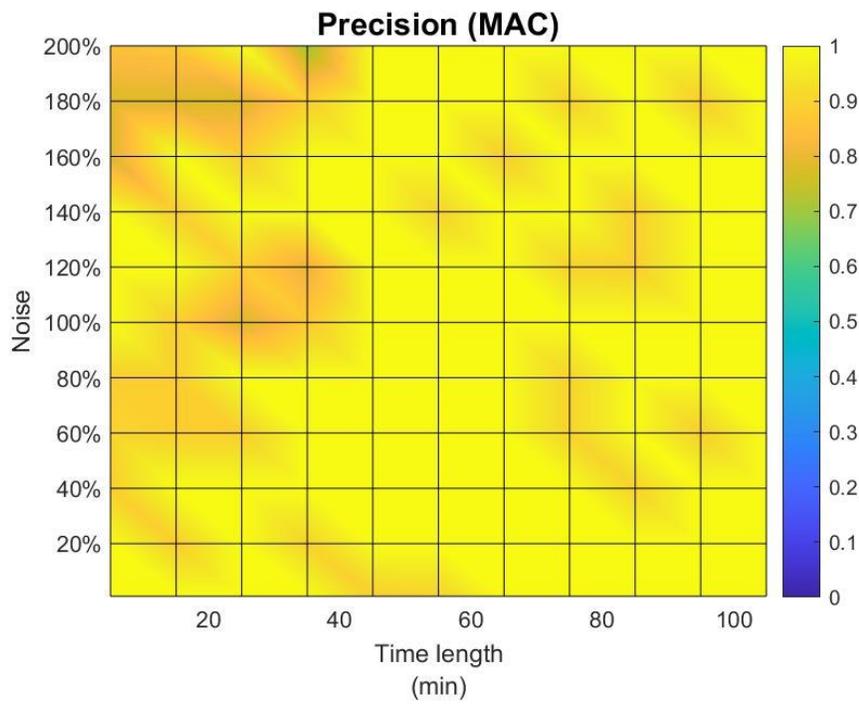


Figure 5-23 Precision in terms of MAC

## Numerical case

---

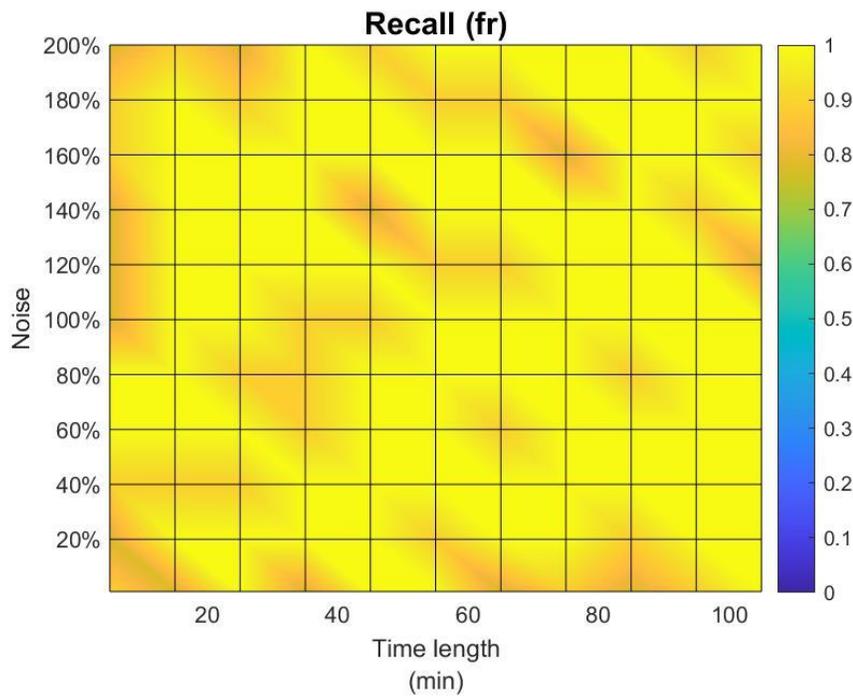


Figure 5-24 Recall in terms of difference of frequencies

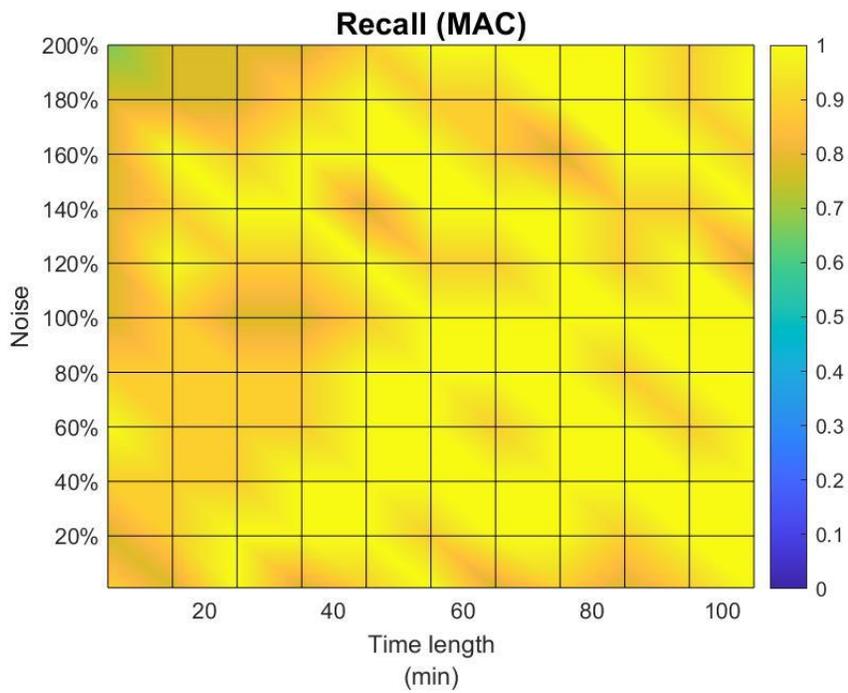


Figure 5-25 Recall in terms of MAC

## Numerical case

---

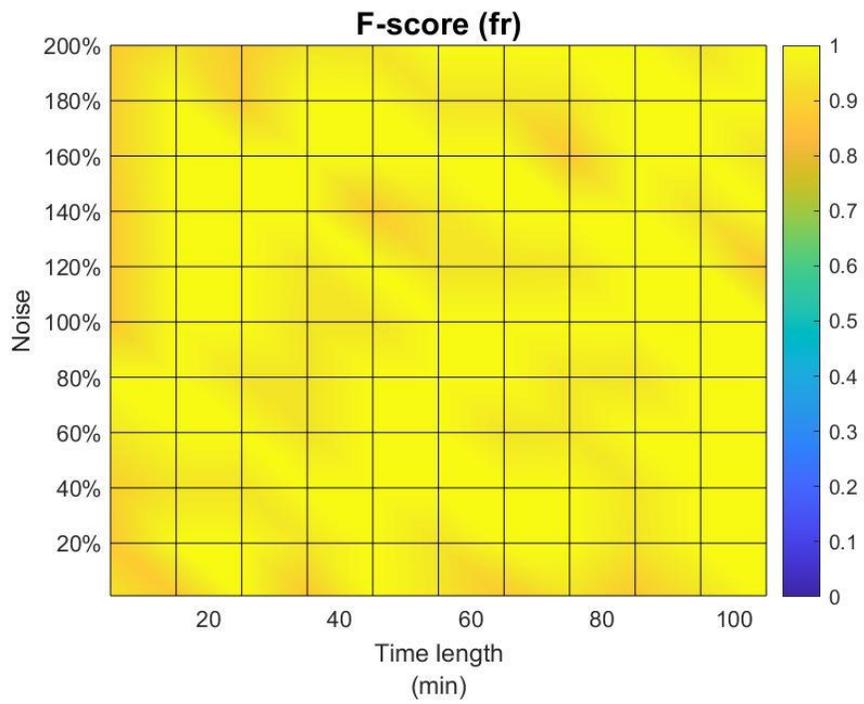


Figure 5-26 F-score in terms of difference of frequencies

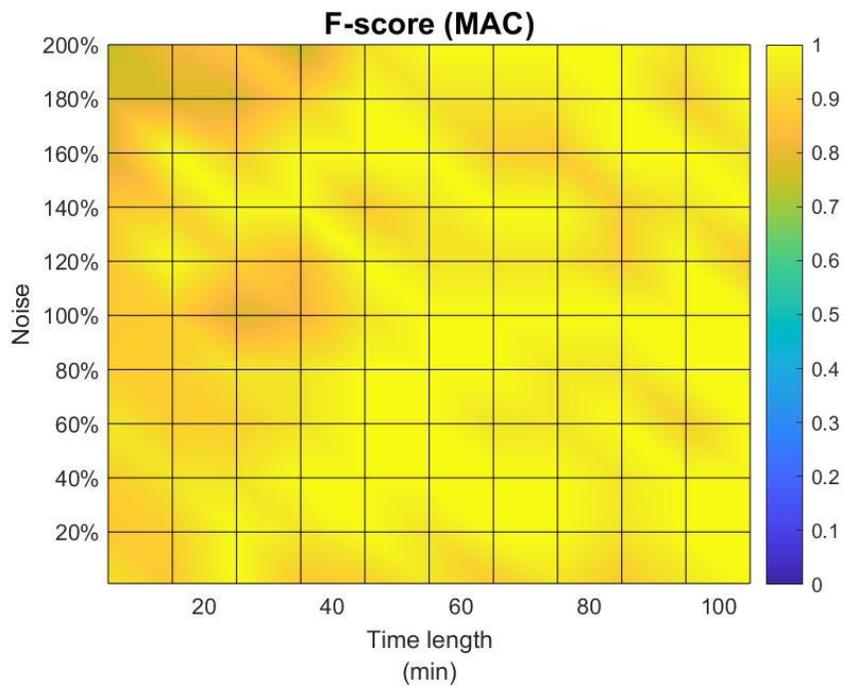


Figure 5-27 F-score in terms of MAC

### 5.3.3 Conclusions

Firstly, the high values of Precision, Recall and F-score confirm the robustness of the proposed AOMA method even with highly noisy signals. The F-score shows values above 0.8 basically in all the cases and close to 1 in most of them, meaning good accuracy of the modal parameter identification.

However, the Recall shows slightly lower values with respect the Precision. This is evidence of the fact that the number of modes identified by the AOMA is less than the number of modes representative of the structure (i.e. 9). Hence, the proposed method may neglect some modes in particular conditions, even though the problem probably occurs in the output signal processing.

The values of the score metrics computed with the frequency identification are generally higher than the values associated with the mode shape identification. Therefore, the mode shapes are identified with a higher level of uncertainty. This flaw can be probably attributed to the identification of the eigenvectors computed by the SSI method and not by the AOMA which select and merge the physical poles automatically.

About the influence of noise and signal duration, the values of scores with the MAC identification appear lower with high noise and low duration. The lower value corresponds to the Recall score computed for the output signal with 200% noise and 5 min duration. In this case 6/9 mode shapes are identified. Contrarily, no trends are recognised in the scores computed with the frequency.

In summary:

- Values of Precision, Recall and F-score close to 1 in most of the cases
- Higher values of Precision with respect to Recall
- Higher values in score metrics computed with the frequency
- Signals with high noise and low duration appear less accurate in terms of mode shape

## Numerical case

---

- There are no significant trends in terms of frequency

# 6 Helicopter blade

In this chapter, the analysis of an experimental case regarding a helicopter blade is presented. This second case study confirms the accuracy and the reliability of the proposed AOMA method.

## 6.1 Description

The study is carried out on an Airbus Helicopter H135 bearingless main rotor blade (BMR) that has been the subject of recent studies at Cranfield University (Depicted in *Figure 6-1*). The cross section of the blade is variable, and its length is 5.1 meters. 9 triaxial piezoelectric PCB 356A45 accelerometers have been applied on the blade. 19 channels in total along the flapwise, the edgewise and the blade axis directions have been recorded. The output signal has been recorded using a standard LabView software at a sampling frequency of 2560 Hz. The measurements have been acquired in two following days at the Aeroelastic laboratory of the Aircraft Integration Research Centre of Cranfield University with controlled room temperature and humidity conditions. The blade has been excited for 600 seconds with an Ambient Vibration Test (AVT) that used an excitation waveform over a frequency of 0-100 Hz. Moreover, the modal parameters have been benchmarked using a Finite Element Model method and the results are compared with the modal parameters identified with the AOMA method.

## Helicopter blade

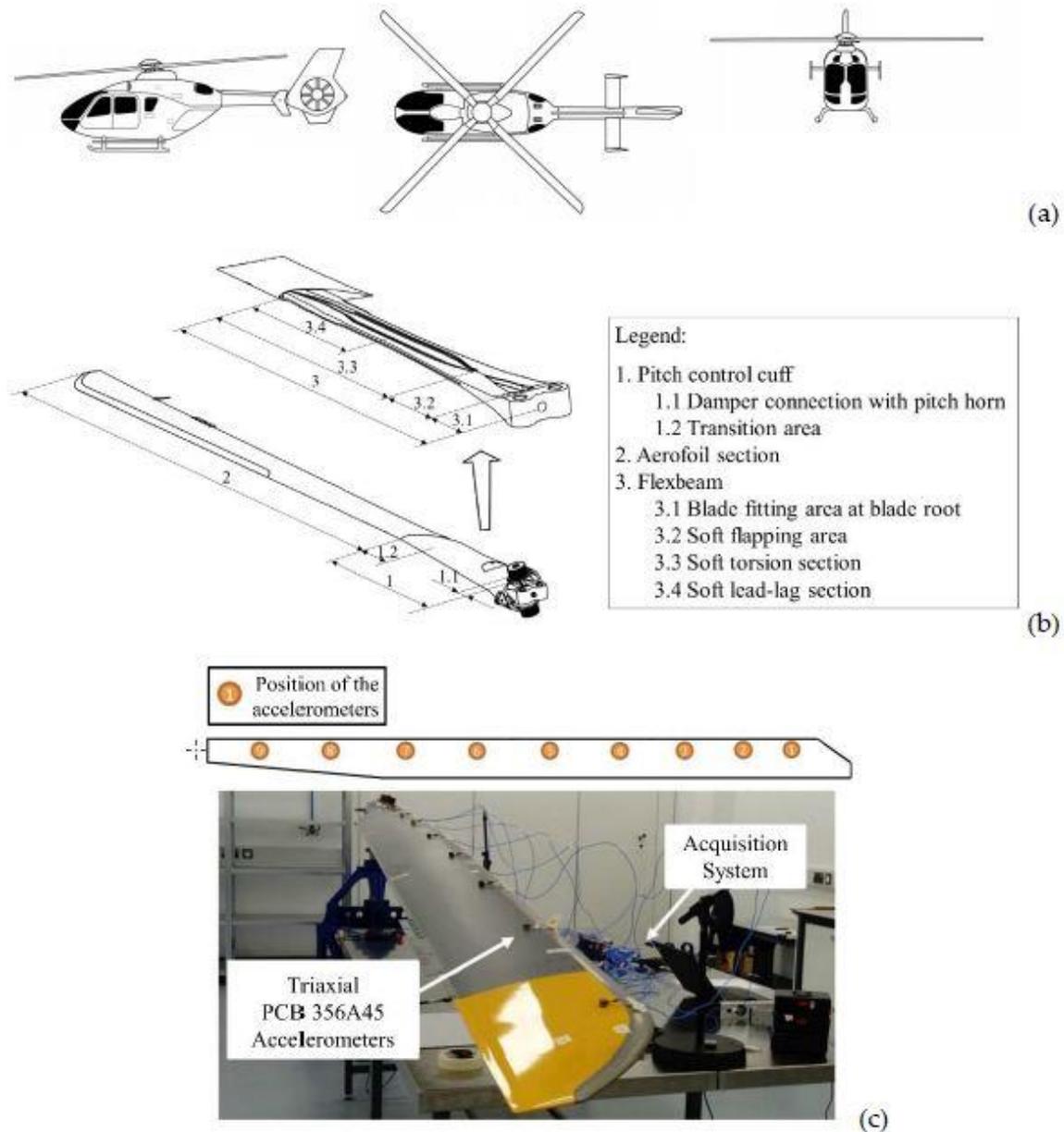


Figure 6-1 (a) Airbus Helicopter H135 general arrangement. (b) main parts of a BMR blade. (c): picture of the investigated BMR blade, acquisition system, and experimental setup

## 6.2 Automated Operational Modal Analysis

In this paragraph, the AOMA method is applied to the presented helicopter blade case study. For conciseness the Hard Validation, the Soft Validation criteria and the

mode shapes identified are omitted and reported in the Appendices. However, the clusters identification and the final modal parameters are described and compared with the FEM results.

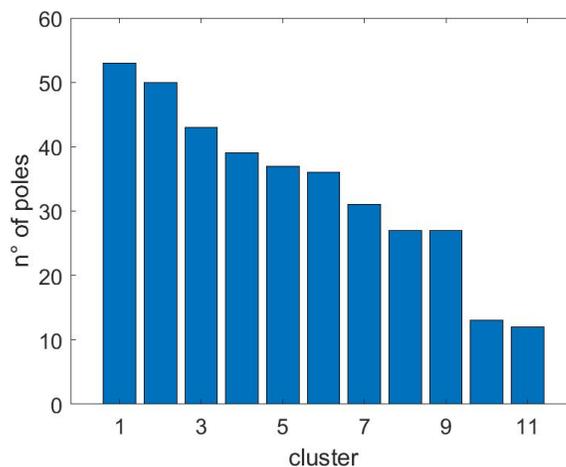
Although the signal has been recorded at 2560 Hz, it is decimated using the Matlab command “Decimate” to decrease the computational cost. Decimate indeed reduces the sample rate of the signal by a specified factor that is chosen equal to 10. Thus, the final sample frequency of the signal becomes equal to 256 Hz not affecting the final modal identification given that the higher expected frequency is about 80 Hz.

The following settings are used in the SSI-COV algorithm:

- Model order ranges between 20 to 130
- Block rows of the Henkel matrix calculated as in eq. 4-2

### 6.2.1 DBSCAN

As in the numerical simulation, the stable poles are grouped in base of their similarity in clusters that are representative of the modal parameters of the system. In this case study, the 11 clusters are obtained merging the 436 stable poles identified by the SVC. Moreover, the dimension of clusters (*Figure 6-2*) highlights the uneven distribution of the poles caused by the accumulation of noise not present in the numerical simulation.



*Figure 6-2 Number of poles in each cluster*

## Helicopter blade

---

The silhouette index of outliers assumes always negative values due to their good separation from clusters. Contrarily, all the stable poles are represented by silhouette values very close to 1.

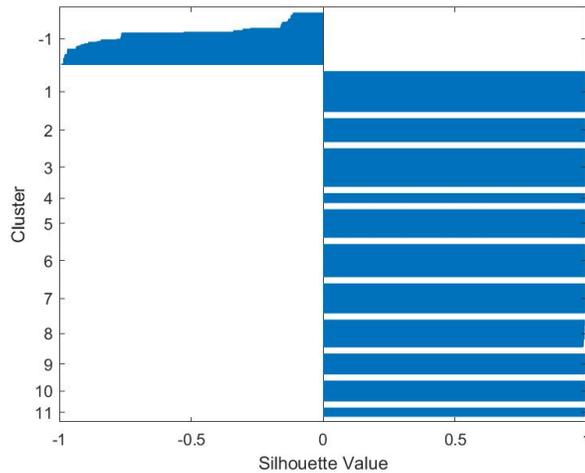


Figure 6-3 Silhouette index

The identified clusters are depicted in the order-frequency-damping ratio diagram with different colours (Figure 6-4). The outliers are represented by cross points.

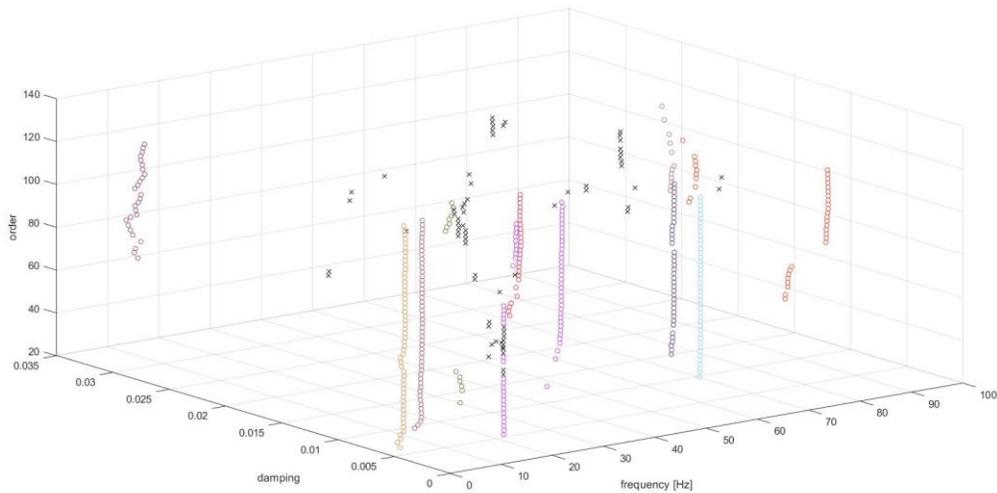


Figure 6-4 Identified clusters after DBSCAN

### 6.2.1.1 MinPts determination

Following the methodology described in **Methodology**, the MinPts parameter used to perform the DBSCAN algorithm is determined by means of the silhouette index. In this case, the best and higher value of the silhouette corresponds to the lower value of the MinPts considered.

Worthy of note is the decreasing trend of the silhouette index increasing MinPts due to the higher number of outliers identified.

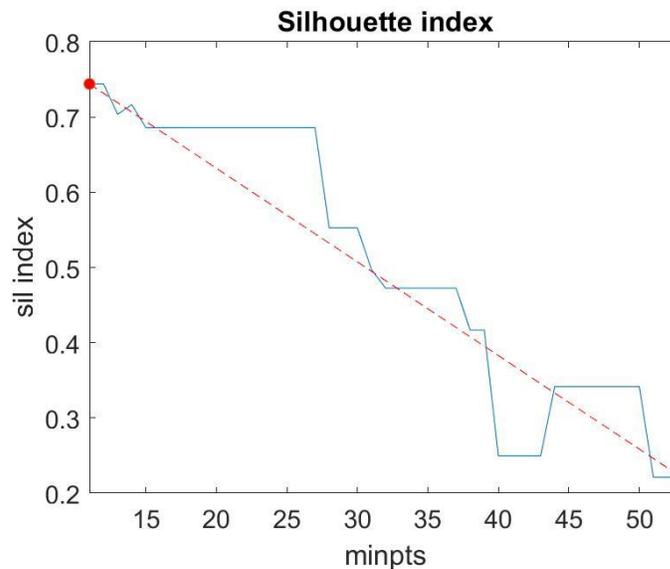


Figure 6-5 MinPts determination

### 6.2.1.2 Epsilon determination

This case study confirms that the heuristic elbow rule method applied to the sorted k-nearest neighbor distance is a robust technique to find the point that precedes the sharp increase of distances. The red dot in *Figure 6-6* corresponds to the k-nn distance used as epsilon parameter to performs the DBSCAN. Its value is 0.0054.

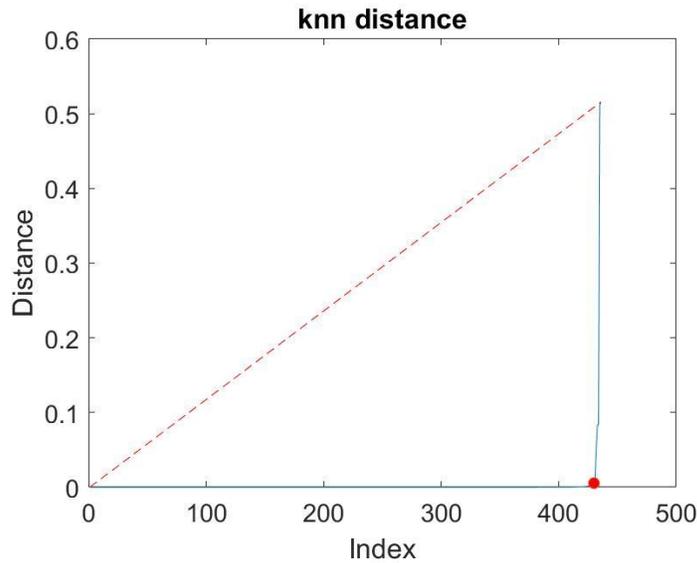


Figure 6-6 Epsilon determination

### 6.2.1.3 Final Modal Parameters

The modal parameters are calculated using the equations described in **Final modal parameter selection**. As already noted in the previous case study, the difference of the calculated frequencies is neglectable (maximum difference between frequencies among the methods ~0.1 Hz)

	Type of mode	Method 1 Identification [Hz]	Method 2 Identification [Hz]	Method 3 Identification [Hz]	Method 4 Identification [Hz]	Method 5 Identification [Hz]
$f_1$	1th Flapping	1.0169	1.0166	1.0170	1.0170	1.0170
$f_2$	1th Lagging	2.6625	2.6619	2.6623	2.6658	2.6629
$f_3$	2th Flapping	5.4324	5.4332	5.4323	5.4323	5.4321
$f_4$	3th Flapping	15.5572	15.5545	15.5519	15.5519	15.5677
$f_5$	2th Lagging	17.9519	17.9713	17.9713	17.9713	17.9809
$f_6$	1th Torsional	28.1241	28.1214	28.1312	28.1312	28.1318
$f_7$	4 <sup>th</sup> Flapping	30.0562	30.0559	30.0565	30.0565	30.0596
$f_8$	1th Spurious	49.1232	49.1239	49.1237	49.1237	49.1226
$f_9$	5 <sup>th</sup> Flapping	51.4349	51.4367	51.4384	51.4384	51.4282
$f_{10}$	6 <sup>th</sup> Flapping	75.7893	75.7961	75.8030	75.8030	75.8933
$f_{11}$	2th Torsional	81.6041	81.6033	81.5923	81.5814	81.6033

Table 6-1 Frequencies identified for the experimental case study

The damping ratios of the identified modes are represented in *Table 6-2*. The scattering of the damping ratios identified in AOMA is a well-known problem. The choice of the identification method indeed slightly influences the identification for the more scattered clusters.

	Type of mode	Method 1 Identification [Hz]	Method 2 Identification [Hz]	Method 3 Identification [Hz]	Method 4 Identification [Hz]	Method 5 Identification [Hz]
$dr_1$	1th Flapping	0.4564	0.4570	0.4609	0.4609	0.4467
$dr_2$	1th Lagging	2.8968	2.8962	2.8755	2.8566	2.8552
$dr_3$	2th Flapping	0.5013	0.5008	0.4934	0.4934	0.4988
$dr_4$	3th Flapping	0.1982	0.2285	0.2334	0.2334	0.1283
$dr_5$	2th Lagging	0.7800	0.7911	0.7911	0.7911	0.8603
$dr_6$	1th Torsional	0.6773	0.6742	0.6627	0.6627	0.6556
$dr_7$	4 <sup>th</sup> Flapping	0.3865	0.3868	0.3801	0.3801	0.3751
$dr_8$	1th Spurious	0.0159	0.0159	0.0157	0.0157	0.0148
$dr_9$	5 <sup>th</sup> Flapping	0.3600	0.3606	0.3545	0.3545	0.3563
$dr_{10}$	6 <sup>th</sup> Flapping	0.5117	0.4788	0.4481	0.4481	0.6830
$dr_{11}$	2th Torsional	1.7679	1.7653	1.7520	1.7300	1.7653

*Table 6-2 Damping ratios identified for the experimental case study*

Besides, the estimated modal parameters are compared to the results of the Mugnaini et al's AOMA method [4] and to the modal parameters calculated with the FEM.

The frequency 56.24 Hz is recognized by the FEM but not by the AOMA. The flaw is due to the choice of the model order or the epsilon determination. The stabilization diagram reported in the **APPENDICES** indeed shows a mode at such frequency. Implementing a higher epsilon probably the mode at 56.24 would have been identified but some spurious mode would have been recognized as physical. The settings implemented in this AOMA algorithm are the right compromise in most of applications. No particular differences between the proposed and the Mugnaini's AOMA methods are highlighted. However, the presented AOMA is even able to recognize one more mode that have been neglected in the Mugnaini's method.

## Helicopter blade

---

The Mugnaini's identification has been carried out setting the range of model order between 20 to 100 and the block rows of the Henkel matrix equals to 6 times the ratio between the maximum model order and the number of channels as suggested by the Authors.

The frequency 49.12 Hz is established as physical, but it's not identified by the FEM.

	Method 1 Identification [Hz]	Mugnaini et al.'s identification [Hz] (difference)	FEM [Hz] (difference)
$f_1$	1.02	1.02 (0.00%)	0.88 <span style="color: red;">(-13.74%)</span>
$f_2$	2.66	--	3.34
$f_3$	5.43	5.43 (0.00%)	4.88 <span style="color: red;">(-10.13%)</span>
$f_4$	15.56	15.55 <span style="color: red;">(-0.06%)</span>	14.98 <span style="color: red;">(-3.73%)</span>
$f_5$	17.95	17.80 <span style="color: red;">(-0.84%)</span>	22.27 <span style="color: red;">(-24.07%)</span>
$f_6$	28.12	--	28.03 <span style="color: red;">(-0.32%)</span>
$f_7$	30.06	30.04 <span style="color: red;">(-0.07%)</span>	29.39 <span style="color: red;">(-2.23%)</span>
$f_8$	49.12	49.16 <span style="color: red;">(-0.08%)</span>	--
$f_9$	51.43	51.51 <span style="color: red;">(-0.16%)</span>	51.15 <span style="color: red;">(-0.54%)</span>
$f_{10}$	--	56.11	56.24
$f_{11}$	75.79	74.93 <span style="color: red;">(-1.13%)</span>	75.12 <span style="color: red;">(-0.88%)</span>
$f_{12}$	81.60	81.67 <span style="color: green;">(0.09%)</span>	81.80 <span style="color: green;">(0.25%)</span>

Table 6-3 Comparison of the identified frequencies with results of other studies

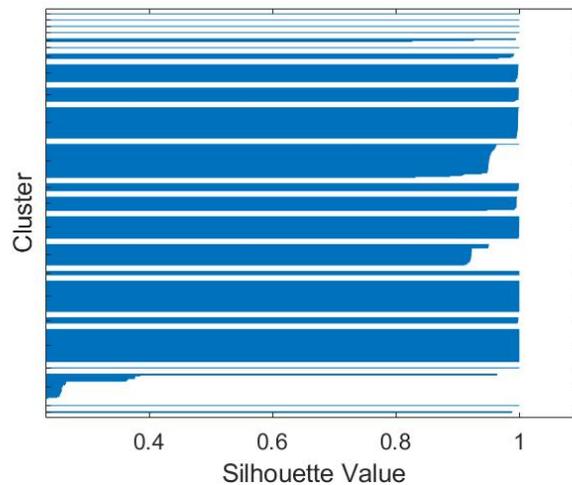
### 6.3 Comparison with Mugnaini et al's results

The previously shown results are compared with the outcomes of the Mugnaini et al's research. It is worth of remembering that the Mugnaini's algorithm clusters the possible physical poles by the use of the hierarchical clustering, identifies the outliers by means of the k-means clustering and sets the block rows of the Henkel matrix differently.

In order to obtain the most similar identification to the FEM results, Mugnaini [4] sets the SSI algorithm with the following parameters:

- Range of model orders with minimum value equal to 20 and maximum value of 100 with intervals of 2.
- Dimension of block rows of the Hankel matrix equals to 6 times the ratio between the maximum model order and the number of channels.

Therefore, the silhouette and the clusters obtained before and after applying the hierarchical and the k-means clustering are depicted in *Figure 6-7*, *Figure 6-8* and *Figure 6-9*. The previous steps are neglected because identical to the ones described in **Automated Operational Modal Analysis**.

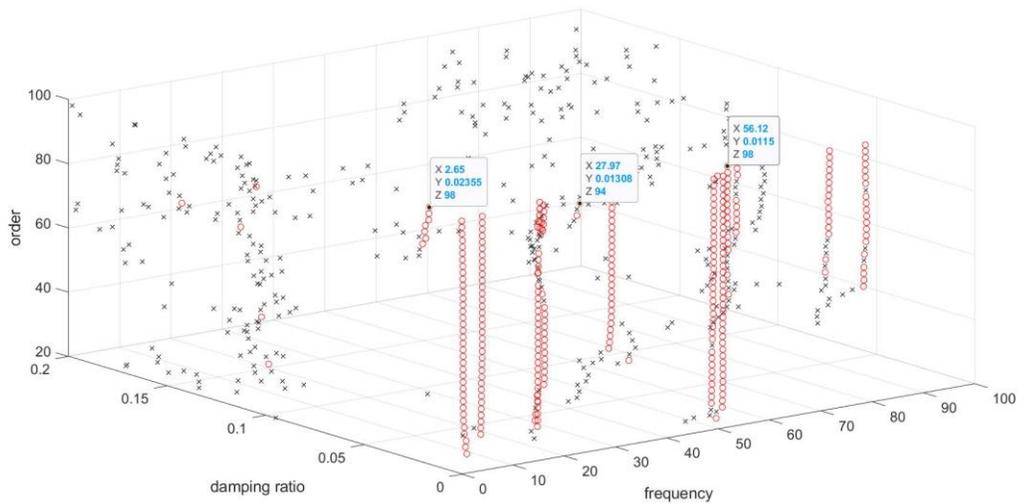


*Figure 6-7 Silhouette value of the clusters identified by Mugnaini*

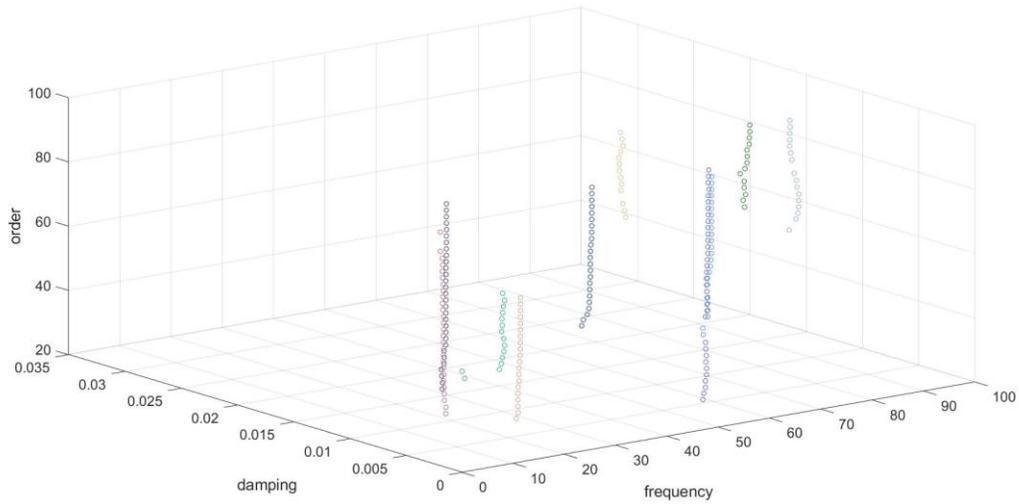
## Helicopter blade

The silhouette shows lower values with respect the one calculated for the DBSCAN clustering. The lack of negative values is due to the remotion of outliers by means of the k-means clustering before performing the silhouette. However, the use of the silhouette index for determining the parameter MinPts led the DBSCAN to obtain higher values that basically correspond to 1 for all the poles.

For understanding how the final modal parameters are obtained, the 3d stabilization diagram acquired before applying the hierarchical clustering and the final cluster representation obtained after implementing both the hierarchical and k-means clustering algorithm are presented in Figure 6-8 and *Figure 6-9*. The pointed-out poles correspond to the modes not identified by both the algorithms.



*Figure 6-8 3D stabilization diagram*



*Figure 6-9 Cluster identification*

Although the high correspondence between the final modal parameters, the Mugnaini's results present the following differences:

- The 2<sup>nd</sup> mode is not identified by the hierarchical clustering. The flaw is probably caused by the overlooking of the physical poles during the clustering phase. Indeed, 7 possible physical poles in correspondence of the 2<sup>nd</sup> mode are disregarded.
- The 6<sup>th</sup> mode is not identified at all. Changing the SSI parameters probably more possible physical poles would have been identified.
- The 10<sup>th</sup> mode is not identified by the proposed AOMA algorithm probably for the selected DBSCAN parameters. Even if the proposed setting does not permit an identical identification to the FEM, it is very good compromise that led to a better identification of Mugnaini's method.

## 7 Z24 bridge

As last modal analysis application, the modal analysis of the Z24 Bridge is considered. The three-span Z24 bridge is located in Switzerland between Bern and Zurich, and it has been subjected of several progressive damage tests in the framework of the SIMCES project in 1997 and 1998.

The AOMA is performed on different output signals recorded during the monitoring and compared with the results of B. Peeters and De Roeck [5] and Reynders et al [2].

### 7.1 Description

The Z24 bridge has been built in 1963 to connect the villages of Koppigen and Utzenstorf overpassing the A1 highway between Zurich and Bern. It was a post-tensioned concrete box girder bridge with a main span of 30 m and 2 side spans of 14 m. The abutments consisted in three pinned concrete columns connected with concrete hinges to the girder, whereas two concrete piers clamped into the girders were situated at the borders of the main span. The bridge has been demolished in 1998 because a new railway adjacent to the Z24 required a larger bridge [25].



*Figure 7-1 Z24 Bridge, Koppigen-Utzenstorf, Switzerland*

## Z24 bridge

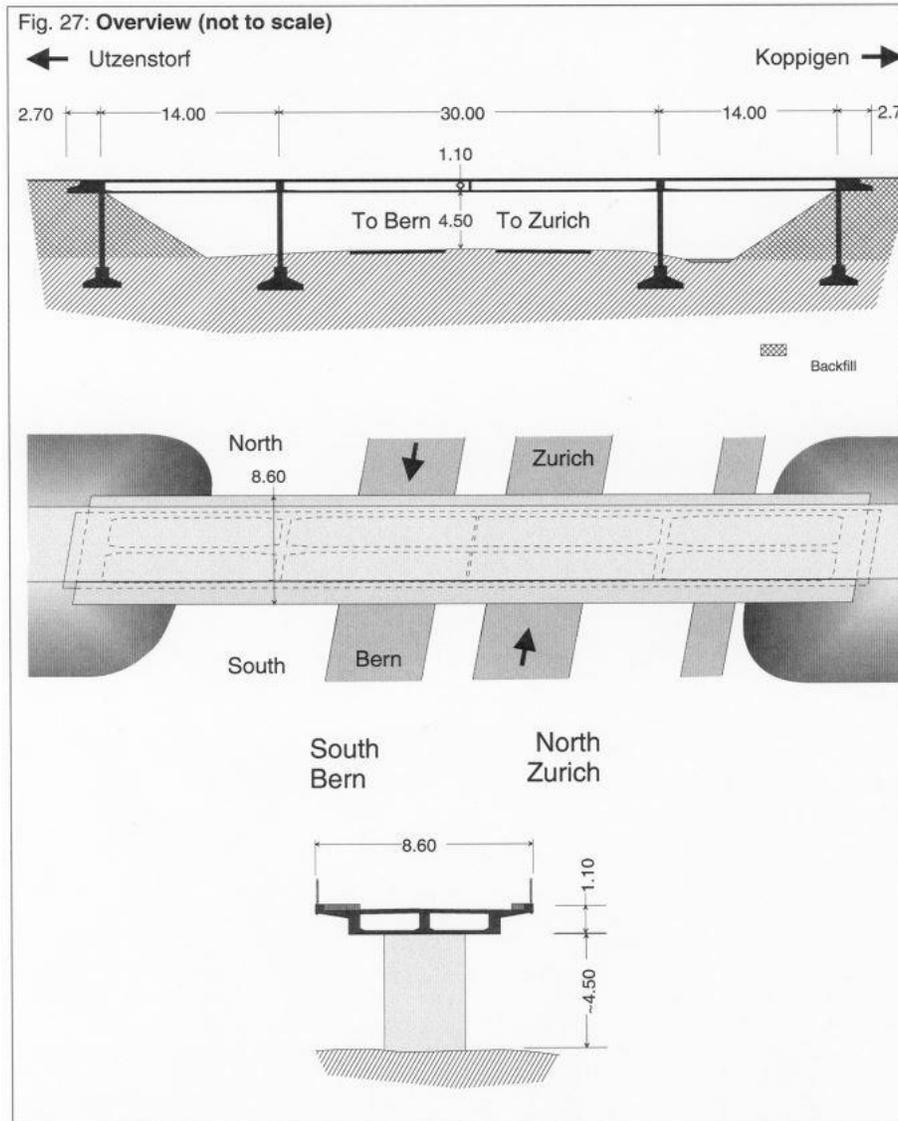
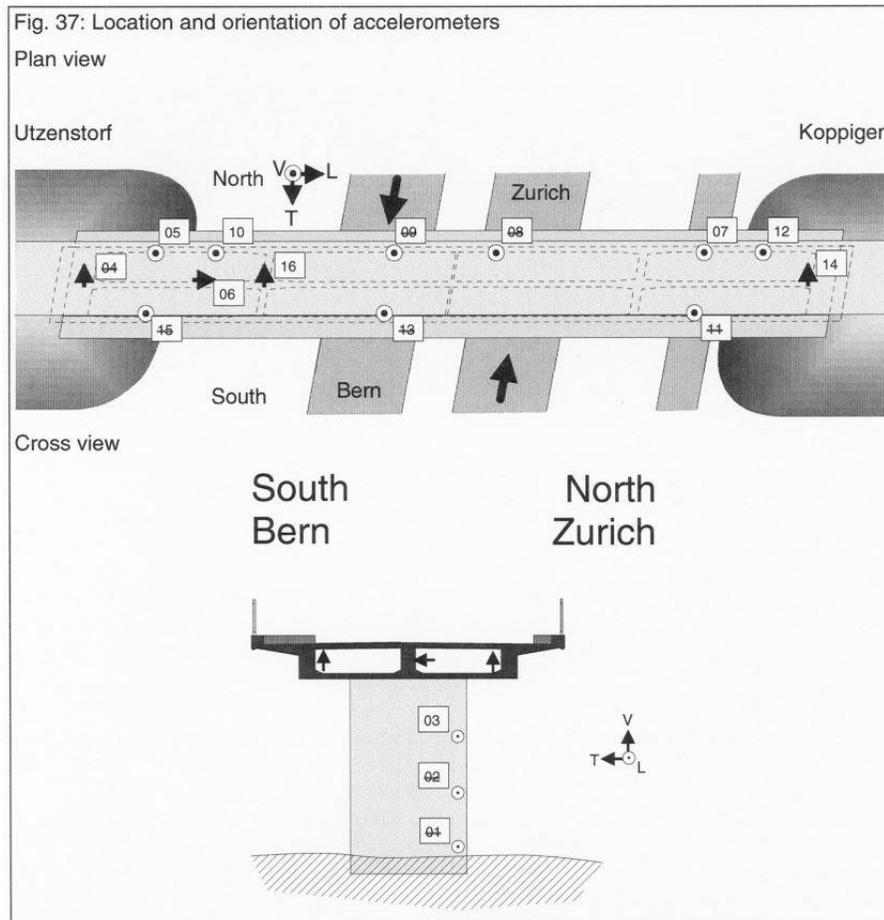


Figure 7-2 Cross-section of the Z24-Bridge

## 7.2 Experimental setup

However, the Z24 bridge was monitored from 11 November 1977 till 11 September 1978 by environmental monitoring system (EMS) measurements and progressive damage test (PDT). The aim of the EMS measurements is to provide both environmental (such as temperature, humidity, wind, etc.) and vibration data. For this purpose, in the Z24 bridge 49 sensors (8 accelerometers) have been totally used

but some of them stopped working during the acquisition. Differently, the PDT sensors acquired only vibration measurements and more than 250 degrees of freedom have been measured in total. The signals obtained from the EMS accelerometers are analysed in this study. The position of the sensors is shown in *Figure 7-3*.



*Figure 7-3 Accelerometers location*

The measurements were repeated every hour, resulting in 24 records per day. Each record lasted about 11 minutes collecting the signals of eight accelerometers. The channels collected 65536 data samples recorded at a sample frequency of 100 Hz.

### 7.3 AOMA application and comparisons

Since the output signals contained an offset, the mean and the trend of the signals have been removed before performing the SSI. The mean is manually removed, whereas the removal of the trend is computed subtracting “polyval” in Matlab from the signals.

The SSI is set as follows:

- Range of the model order between 10-160
- Block rows of the Henkel matrix set as in eq. (4-2)

Given that the modal analysis studies performed on the Z24 bridge regard only the Operational Modal Analysis, the modal parameters identified with the presented AOMA are compared with the results of other research. Since only 8 accelerometers placed on the side span have been employed in the EMS study, the mode shapes are not identified and only the eigenfrequencies are investigated.

The first comparison is made considering the results of De Roeck’s study [1005] that investigated the influence of the environmental effects on the identified eigenfrequencies.

The signals analyzed with the AOMA are obtained from the following acquisitions:

- Acquisition 300, 24/11/1997, time 01:00, day 13
- Acquisition 605, 06/12/1997, time 18:00, day 25
- Acquisition 1827, 29/01/1998, time 09:00, day 79
- Acquisition 3006, 25/03/1998, time 00:00, day 134
- Acquisition 4048, 25/06/1998, time 11:00, day 225

Where the aforementioned descriptions are organized as follows:

- Number of acquisition, date of acquisition, time of acquisition and day of acquisition

## Z24 bridge

For conciseness only the stabilization and the order-frequency-damping ratio diagrams are depicted.

### 1. Acquisition 300

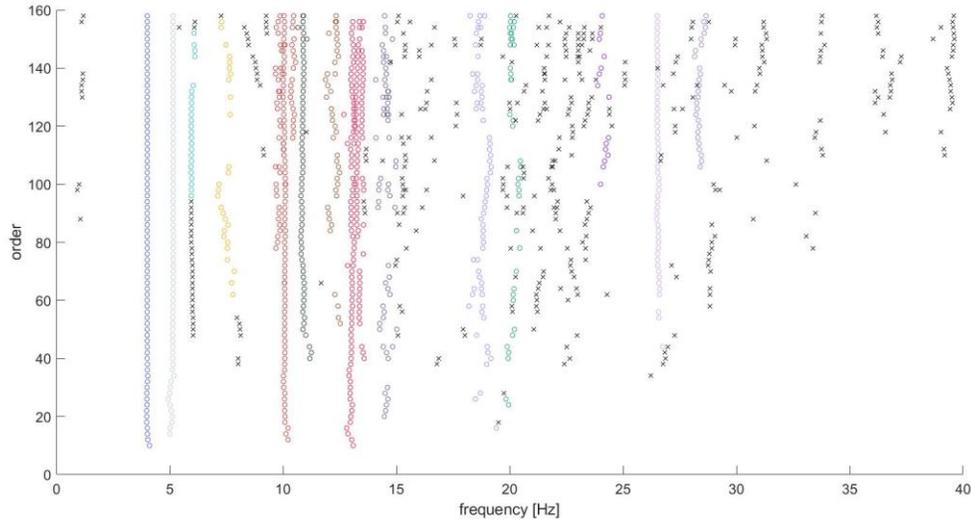


Figure 7-4 Acquisition 300 2D clusters identification

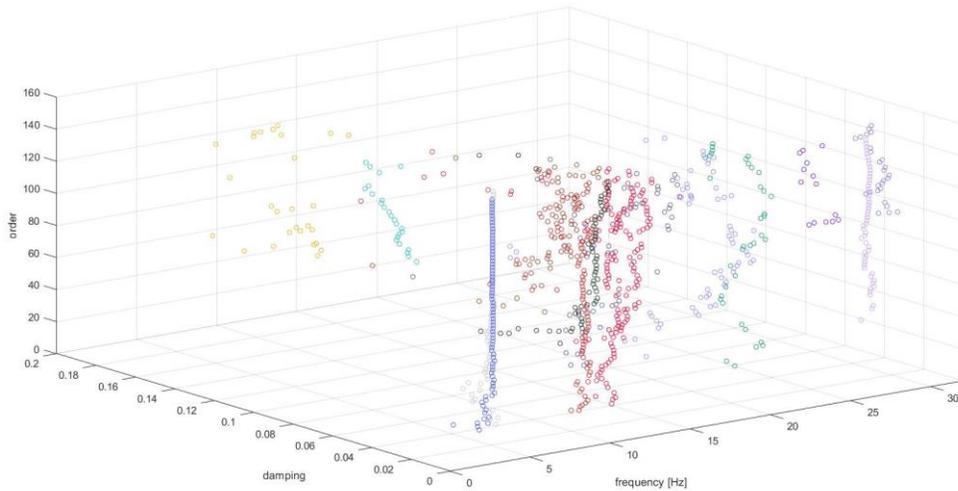


Figure 7-5 Acquisition 300 3D clusters identification

## 2. Acquisition 605

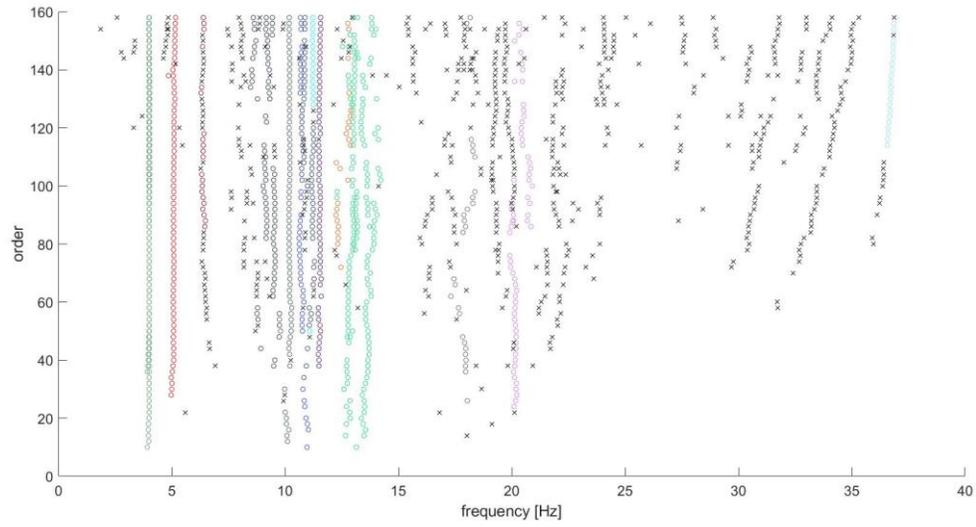


Figure 7-6 Acquisition 605 2D clusters identification

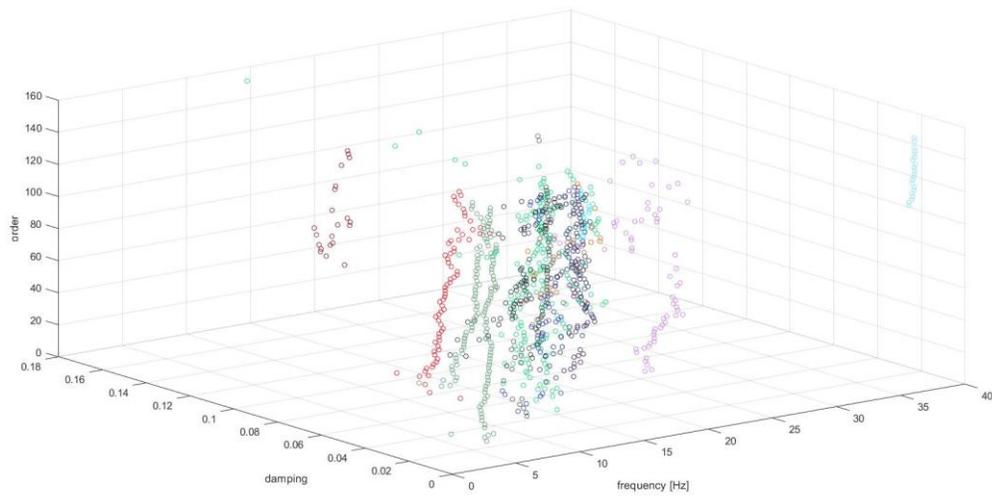


Figure 7-7 Acquisition 605 3D clusters identification

### 3. Acquisition 1827

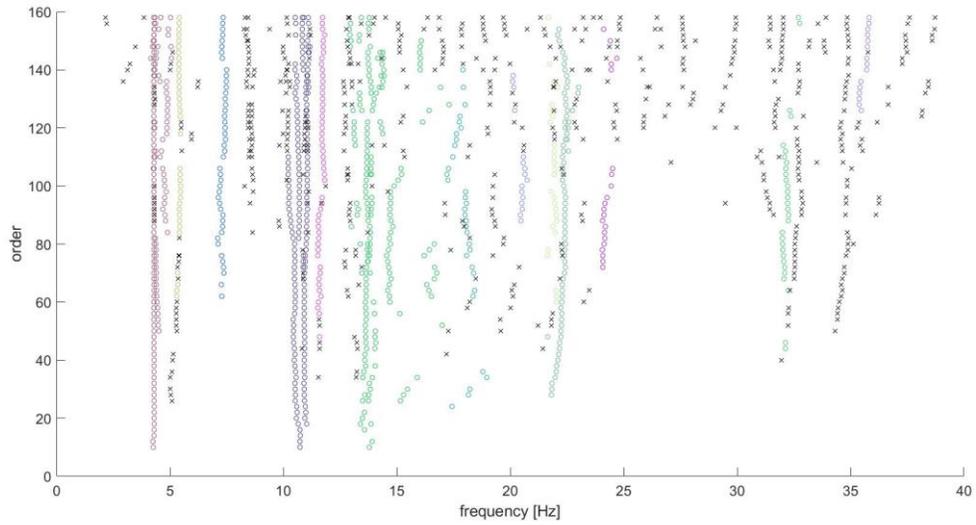


Figure 7-8 Acquisition 1827 2D clusters identification

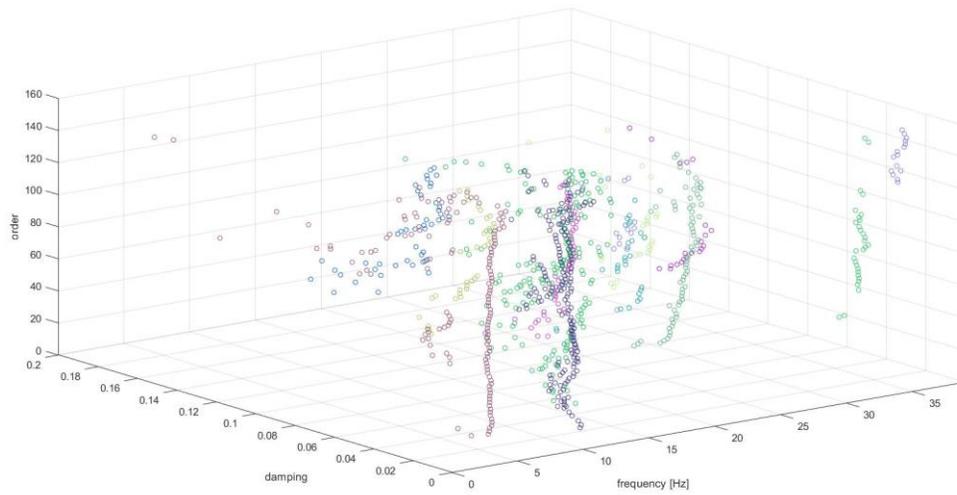


Figure 7-9 Acquisition 1827 3D clusters identification

4. Acquisition 3006

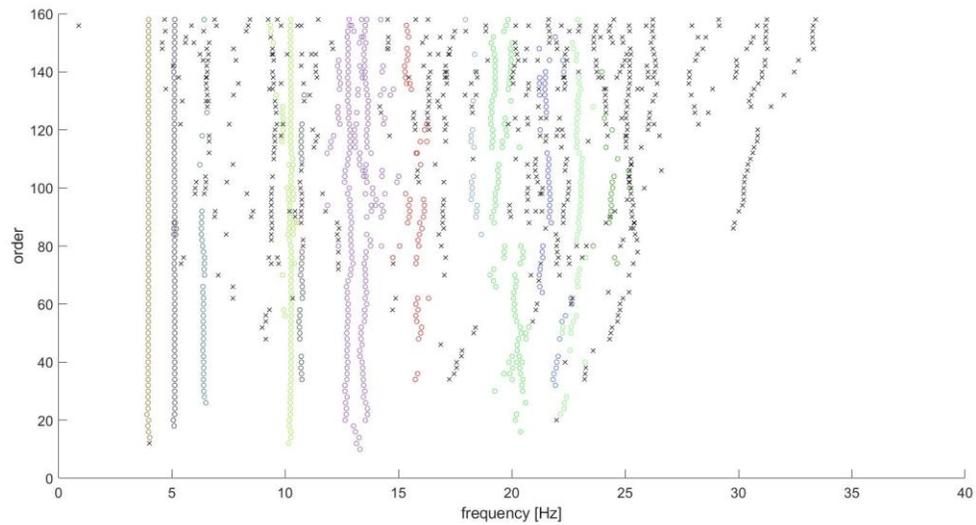


Figure 7-10 Acquisition 3006 2D clusters identification

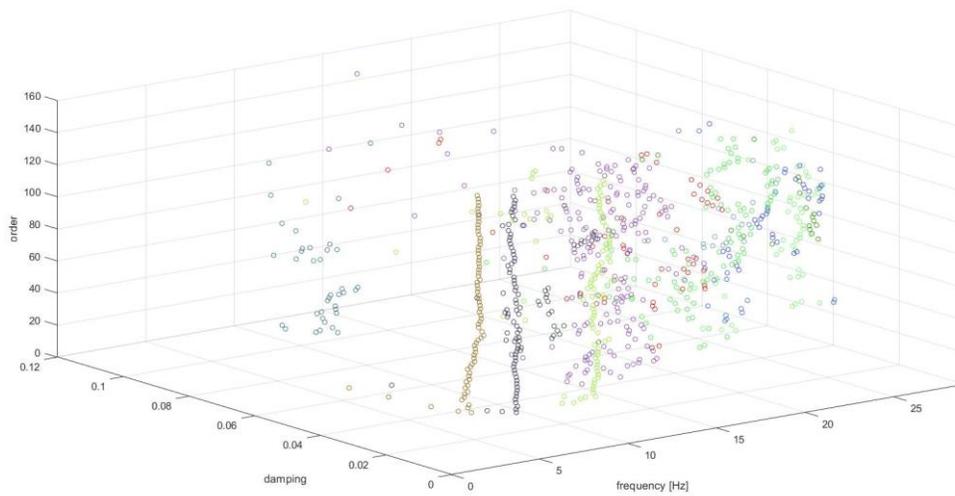


Figure 7-11 Acquisition 3006 3D clusters identification

5. Acquisition 4048

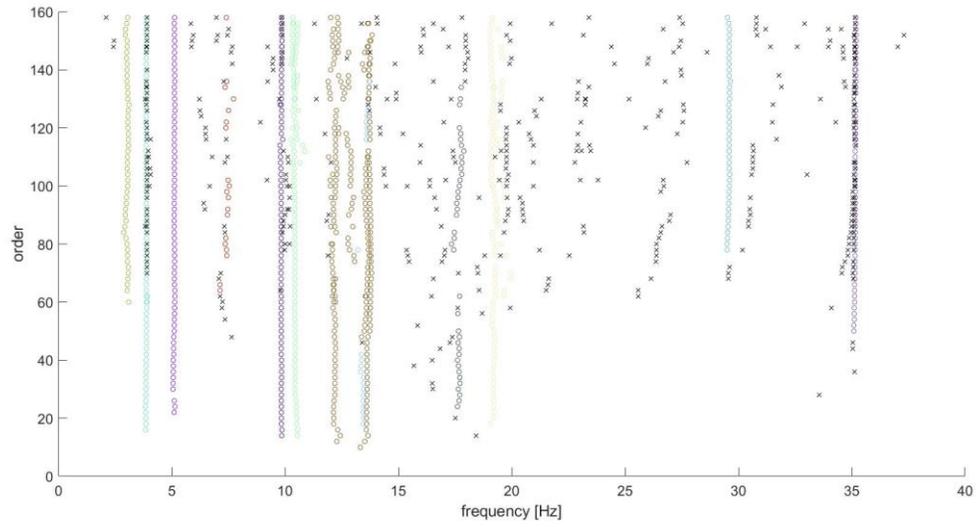


Figure 7-12 Acquisition 4048 2D clusters identification

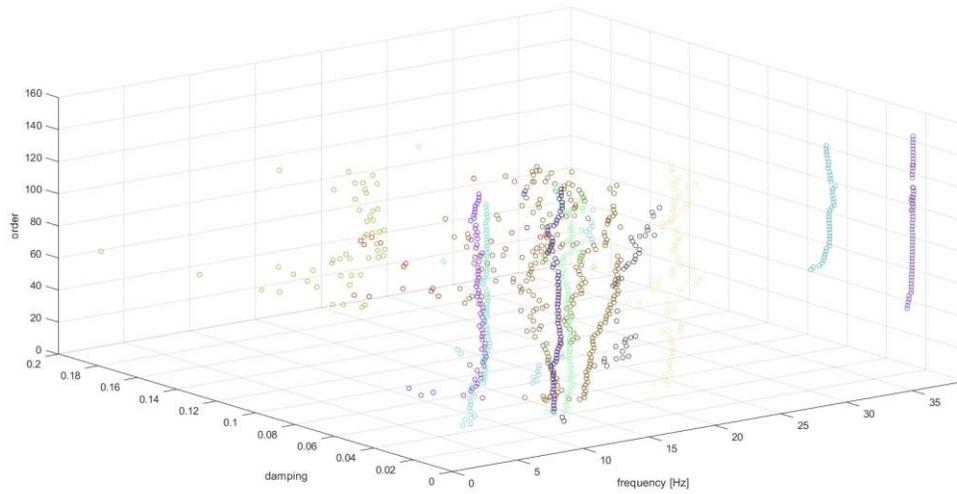
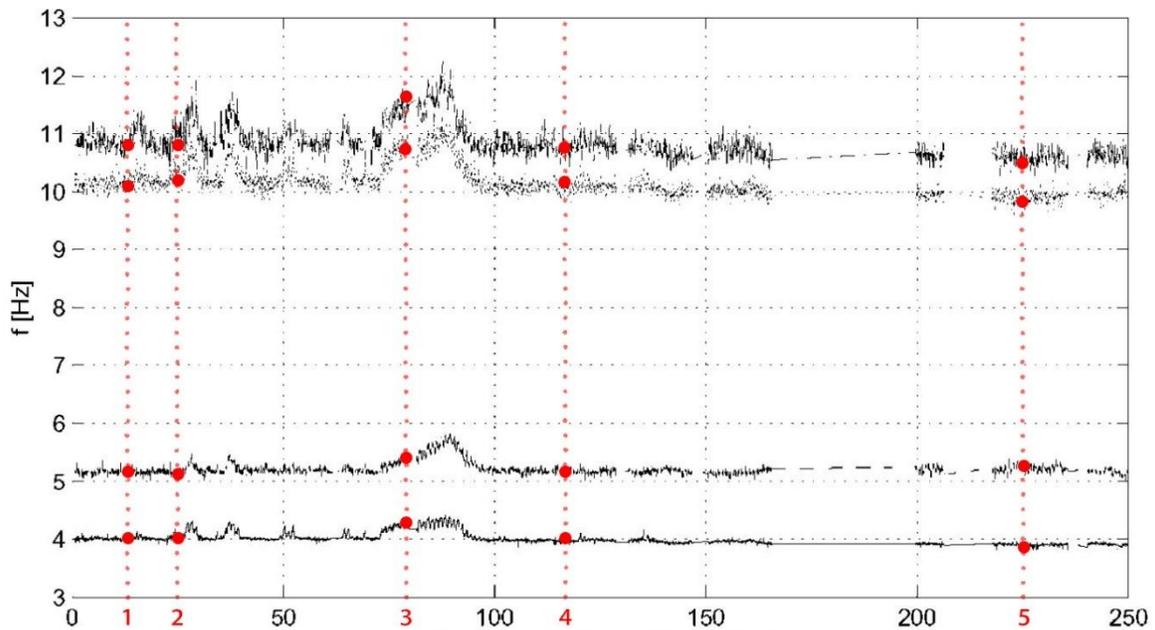


Figure 7-13 Acquisition 4048 3D clusters identification

From preliminary modal surveys [18], we know that four modes are always identified in the range 0-12 Hz. The frequencies of these modes have been plotted as function of time by De Roeck from the 1<sup>st</sup> till the 250<sup>th</sup> day of monitoring and compared to the frequencies identified with the described AOMA algorithm. During the investigation different weather conditions were considered that even caused structural stiffening (from 20 January to 13 February a prolonged acute freezing period happened). In *Figure 7-14* the black dots are the identifications collected by De Roeck, whereas the red dots are representative of the eigenfrequencies obtained with the presented AOMA.

Clearly, all the four eigenfrequencies are correctly identified with the AOMA method in all weather conditions, proving its encouraging performances.



*Figure 7-14* The frequencies of the first 4 modes have been plotted as function of time by De Roeck from the first till the 250<sup>th</sup> day of monitoring and compared to the frequencies identified with the described AOMA algorithm.

The identified eigenfrequencies are secondly compared with the Reynder's identifications [2]. Even if the acquisitions used are never mentioned, we supposed that the investigation has been made during the Short-term progressive damage test.

This assumption is made because they used the PDT sensors that have been employed only in the last month of monitoring.

Reynders analyzed 9 different setups identifying 6 common modes. In each setup several modes have been identified but only the modes in common have indeed been considered as representative of the structure.

The signals analyzed with the AOMA are obtained from the following acquisitions:

- Acquisition 4906, 09/08/1998, time 12:00
- Acquisition 5403, 30/08/1998, time 08:00
- Acquisition 5595, 07/08/1998, time 23:00

Since, more than 6 modes have been identified in each signal, only the modes in common with the Reynders' identification are reported (*Table 7-1*). Considering that Reynders didn't specify the acquisitions that they used, the results show strong similarity for exception of the fifth mode that is not recognized in acquisition 4906 and in acquisition 5595.

## Z24 bridge

	Acquisition 4906 [Hz] (difference)	Acquisition 5403 [Hz] (difference)	Acquisition 5595 [Hz] (difference)	Average difference	Reynders et al.'s identification [Hz]
$f_1$	3.90 (+1.04%)	3.89 (+0.77%)	3.81 (-1.29%)	+0.17%	3.86
$f_2$	4.92 (+0.40%)	4.75 (-3.06%)	4.68 (-4.49%)	-2.38%	4.90
$f_3$	9.82 (+0.61%)	9.88 (+1.23%)	9.70 (-0.61%)	+1.23%	9.76
$f_4$	10.25 (-0.48%)	10.34 (+0.39%)	10.23 (-0.68%)	-0.25%	10.30
$f_5$	--	12.34 (-0.56%)	--	-0.56%	12.41
$f_6$	12.98 (-1.81%)	13.47 (+1.89%)	13.41 (+1.44%)	0.51%	13.22

*Table 7-1 Comparison of the identified frequencies with the Reynders' results*

In the stabilization diagram of the acquisition 4906 there is a clear vertical line in correspondence of the 5<sup>th</sup> mode, but it is merged with poles with higher frequency. Ignoring the mode shapes, it's difficult to understand why the algorithm groups poles with such difference in frequency. Probably their eigenvectors are so similar that they are grouped in the same cluster.

## Z24 bridge

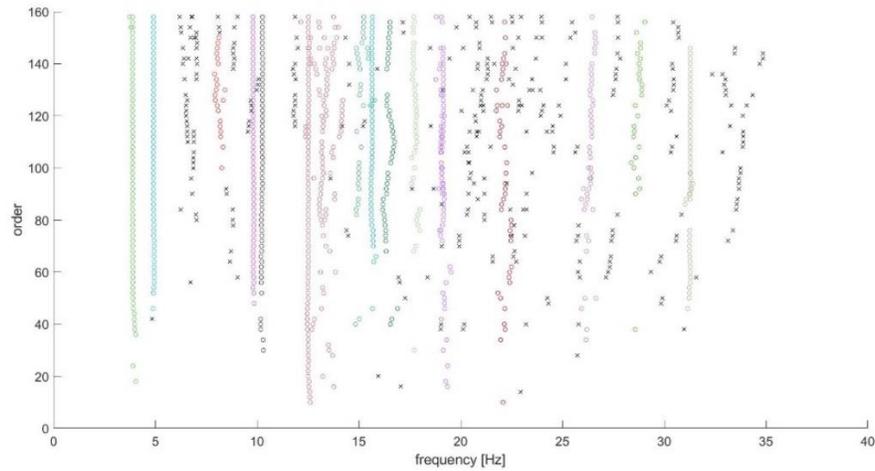


Figure 7-15 Identified clusters of acquisition 4906

Contrarily, analyzing the stabilization diagram of the acquisition 5595 obtained after applying the Hard Validation Criteria, it's clear the 5<sup>th</sup> mode is not identified by the SSI. Therefore, the problem doesn't regard the AOMA algorithm.

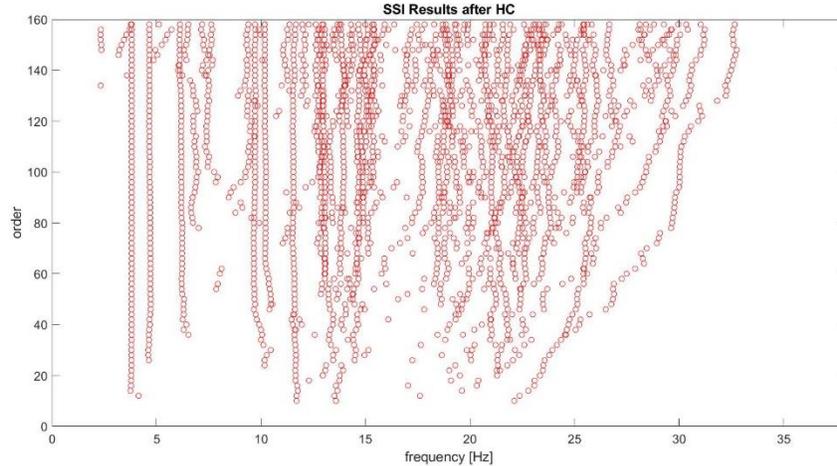


Figure 7-16 Stabilization diagram of acquisition 5595 after HVC

Given the uncertainty of which acquisitions have been used by Reynders and the employment of different sensors to collect the signal, the acquisition 4906, 5403 and 5595 have been analyzed by the Mugnaini's method as well. The results are compared in *Table 7-2*.

## Z24 bridge

Natural frequencies [Hz]				
Blue: identification with the proposed AOMA				
Red: Mugnaini's identification				
	Acquisition 4906 [Hz]	Acquisition 5403 [Hz]	Acquisition 5595 [Hz]	Reynders et al.'s identification [Hz]
$f_1$	3.90 3.89	3.89 3.89	3.81 3.81	3.86
$f_2$	4.92 4.92	4.75 4.73	4.68 4.68	4.90
$f_3$	9.82 9.82	9.88 9.88	9.70 9.71	9.76
$f_4$	10.25 10.25	10.34 10.40	10.23 10.20	10.30
$f_5$	-- 12.52	12.34 12.34	-- --	12.41
$f_6$	12.98 --	13.47 13.44	13.41 13.01	13.22

*Table 7-2 Comparison of results obtained with different AOMA methods*

The first four modes that, as already mentioned, have been always identified in preliminary modal surveys [18] are commonly identified by all the 3 AOMA methods. Contrarily, the 5<sup>th</sup> and the 6<sup>th</sup> modes have not been identified at all. Actually, from the frequency 12 Hz would not be easy identify the vertical columns of stable poles even with the visual inspection of the stabilization diagram by the user (*Figure 7-15* and *Figure 7-16*), indeed the presence of so close poles makes the detection of vertical lines complicated.

# 8 Conclusions

## 8.1 Summary and conclusions

In this thesis a novel multi-stage clustering approach for automated operational modal analysis is presented. It analyses the candidate modes identified by the stochastic subspace identification method by means of hard validation criteria, soft validation criteria and the density-based spatial clustering of applications with noise.

Firstly, an historical background and a brief introduction of Structural Health Monitoring and operational modal analysis are explained in order to lay the research groundwork. The methodologies, the key concepts and the processes used in other similar AOMA methods are successively provided and compared. Finally, the methodology used in the proposed AOMA algorithm and its application in numerical, experimental, and really existed case studies are reported.

The proposed AOMA method relies on the parameters recognized by the SSI that analyses output-only signals collected during operating conditions. The AOMA algorithm is completely automatic since no parameters or thresholds have to be provided by the user. Differently to other AOMA methods, the novelty is the implementation of the DBSCAN clustering algorithm for grouping the physical modes in different clusters based on their similarity. Differently to the reference AOMA methods (Reynders et al' and Mugnanini et al's methods) the DBSCAN allows to identify the outliers and to cluster the modes in only one step, indeed the hierarchical and the k-means clustering algorithms are implemented in the other cases. Usually, DBSCAN needs the definition of two parameters: the epsilon and the minpts. In the proposed algorithm they are provided automatically by the use of a cluster validation criteria and a heuristic method. Moreover, the final modal parameters representative of each cluster are identified applying five different statistical approaches.

The performance and the functionality of the proposed AOMA algorithm have been tested in three different cases that proved its robustness and well-conditioning.

For the numerical case, the frequencies, damping ratios and mode shapes identified have been compared to the results obtained with the eigenproblem showing complete or very high similarity. Furthermore, the signal length and the noise added to the signal have been varied in order to evaluate their influence on the results. The quality of the clustering has been quantified by the use of precision, recall and F1 score that assumed values closed to 1 in basically all of the cases. The only trend recognised is for short signal durations and high noise, in these conditions the recall computed with the MAC showed values around 0.7.

The modal parameters obtained in the experimental case regarding a helicopter blade have been compared with the results of the FEM study and of Mugnaini et al's AOMA method. The presented AOMA have been able to recognize one more mode with respect the Mugnaini et al's method but not all the mode identified with the FEM. Indeed, the tenth mode has been missed.

The final case study regarded the analysis of the really existed Z24 bridge. The analysed output signals have been collected during the monitoring of its last year of life during which several progressive damage tests have been performed. The modal parameters identified have been compared with the results of B. Peeters and De Roeck and Reynders et al's studies. Regarding the first study on which the output signals analysed were known, all the first four modes have been correctly identified in all weather conditions proving the encouraging performance of the algorithm. Differently regarding the Reynders et al's study on which the output signals were unknown, one of the six modes is not always been correctly identified probably for the DBSCAN parameter selection.

### **8.2 Future perspectives**

Future research should focus on the automatization of the DBSCAN and apply the algorithm to further case studies. The determination of epsilon could be made relying

## Conclusions

---

not only to the silhouette index but considering also other evaluation metrics. The selection of MinPts should be investigated as well. Even if the identification of the modal parameters was similar or even better of other AOMA methods, a more appropriated selection of the DBSCAN parameters could enhance the performance of the identification in particular situations.

# References

1. C. Rainieri and G. Fabbrocino, *Influence of model order and number of block rows on accuracy and precision of modal parameter estimates in stochastic subspace identification*, International Journal of Lifecycle Performance Engineering 1(4):317, 2014.
2. E. Reynders, J. Houbrechts, G. de Roeck, *Fully automated (operational) modal analysis*, Mech. Syst. Signal Process. 29 (2012) 228–250.
3. E. Neu, F. Janser, A.A. Khatibi, A.C. Orifici, *Fully Automated Operational Modal Analysis using multi-stage clustering*, Mech. Syst. Signal Process. 84 (2017) 308–323, 2017.
4. V. Mugnaini, L. Zanotti Fragonara, M.Civera, *A machine learning approach for automatic operational modal analysis*, Politecnico di Torino, 2019.
5. Bart Peeters and Guido De Roeck, *One-year monitoring of the Z24-Bridge: environmental effects versus damage events*, Earthquake Engng Struct. Dyn. 2001; 30:149–171, 2001.
6. W. Heylen, S. Lammens, P. Sas, *Modal Analysis Theory and Testing*, n.d
7. R.J. Allemang, D.L. Brown, *A correlation coefficient for modal vector analysis*, in: Proc. 1st Int. Modal Anal. Conf. (IMAC 1982), 1982: pp. 110–116.
8. M. Pastora, M. Bindaa, T. Harparik, *Modal Assurance Criterion*, Procedia Engineering 48 543 – 548, 2012.
9. Lin-Jun Lu, Hua-Fei Zhou, Yi-Qing Ni, Fei Dai, *Output-only modal analysis for non-synchronous data using stochastic sub-space identification*, Engineering Structures 230 (2021) 111702, 2021.
10. Lloyd, P. Stuart, *Least square quantization in PCM*, Bell Telephone Laboratories Paper, 1957.
11. A.K. Jain, R.P.W. Duin, J. Mao, *Statistical pattern recognition: a review*, IEEE Trans. Pattern Anal. Machine Intelligence, 22, 4–37, 2000.

## References

---

12. T.M Kodinariya, P.R. Makwana, *Review of determining of cluster in k-means clustering*, International journal of advance research in computer science and management studies, Volume 1, Issue 6, 2013.
13. S.C. Johnson, *Hierarchical clustering schemes*, PSYCHOMETRIKA-VOL 32, NO. 3, 1967.
14. B. A. Jamous, R. Fa, A. K Nandi, *Hierarchical clustering*, Integrative Cluster Analysis in Bioinformatics, 2015.
15. M.Ester, H.Kriegel, J.Sander, X. Xu, *A density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise*, Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96). AAAI Press. pp. 226–231, 1996.
16. K. Liu, *Modal parameter estimation using the state space method*, J. Sound Vibr., 197 (1996), 387–402, 1996.
17. Zhi Li, Jiyang Fu, Qisheng Liang, Huajian Mao and Yuncheng He, *Modal identification of civil structures via covariance-driven stochastic subspace method*, Mathematical Biosciences and Engineering Volume 16, Issue 5, 5709–5728, 2019.
18. E. Reynders and G. De Roeck, *Reference-based combined deterministic-stochastic subspace identification for experimental and operational modal analysis*, Mechanical Systems and Signal Processing, Vol. 22, No. 3, pp.617–637, 2008.
19. P. Van Overschee and B. De Moor, *Subspace identification for linear systems: Theory, Implementation, Applications*, Kluwer Academic Publishers, 1996.
20. George E. P. Box, D. R. Cox, *An analysis of transformations*, Journal of the Royal Statistical Society, Series B. 26 (2): 211–252, 1964.
21. Peter J. ROUSSEEUW, *Silhouettes: a graphical aid to the interpretation and validation of cluster analysis*, Journal of Computational and Applied Mathematics 20 (1987) 53-65 North-Holland, 1987.

## References

---

22. F. Magalhães, Á Cunha, E. Caetano, *Online automatic identification of the modal parameters of a long span arch bridge*, Mech. Syst. Signal Process. 23(2) (2009) 316–329.
23. J. Schwochow, G. Jelcic, *Automatic operational modal analysis for aeroelastic applications*, in: Proceedings of the 6th International Operational Modal Analysis Conference, Gijón, Spain, 2015.
24. C. Ye, X. Zhao, *Automated Operation Modal Analysis Based on DBSCAN Clustering*, International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), 2020
25. J. Maeck and G. De Roeck, *Description of Z24 benchmark*. Mechanical Systems and Signal Processing, 17(1):127-131, 2003.
26. Marco Sansoni and Fabio Vandin, *On the use of Silhouette for cost based clustering*, Università degli Studi di Padova, Department of Ingegneria dell'Informazione, Master Thesis, 2019.
27. Sara Prone, Paolo Garza and Eliana Pastor, *A Big Data Solution for Silhouette Computation*, Politecnico di Torino, Department of Computer Engineering, Master Thesis, 2019.
28. Nadia Rahmah and Imas Sukaesih Sitanggan, *Determination of Optimal Epsilon (Eps) Value on DBSCAN Algorithm to Clustering Data on Peatland Hotspots in Sumatra*, IOP Conference Series: Earth and Environmental Science, 20.

# APPENDICES

## Helicopter Blade: Hard Validation Criteria

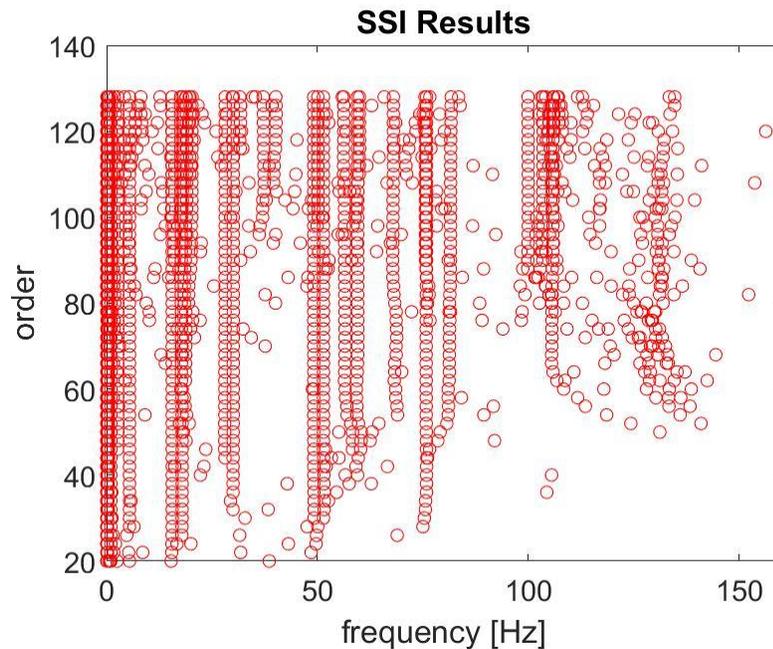
The HVC is performed as described in **Methodology**.

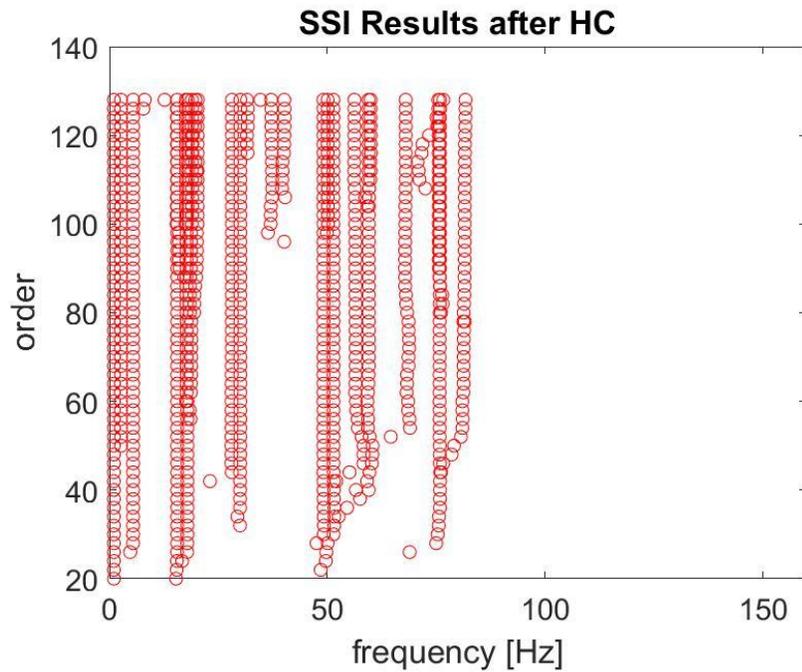
In summary, the cleansing out of the candidate poles obtained by the SSI is performed applying the following criteria:

$$0\% \leq \xi_j \leq 20\%$$
$$Re(\lambda_j) \geq 0, \quad Im(\lambda_j) \neq 0$$

The poles that don't satisfy these criteria are labelled as certainty mathematical and discarded, instead the other poles are labelled as possible physical.

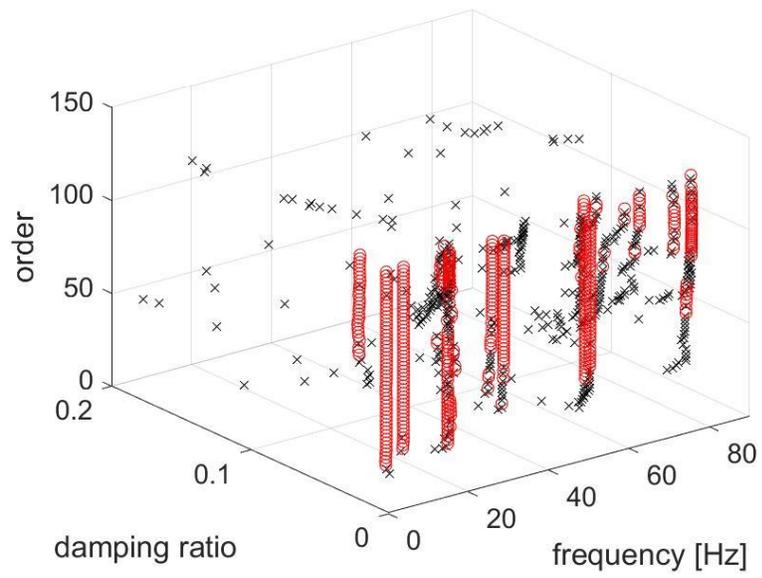
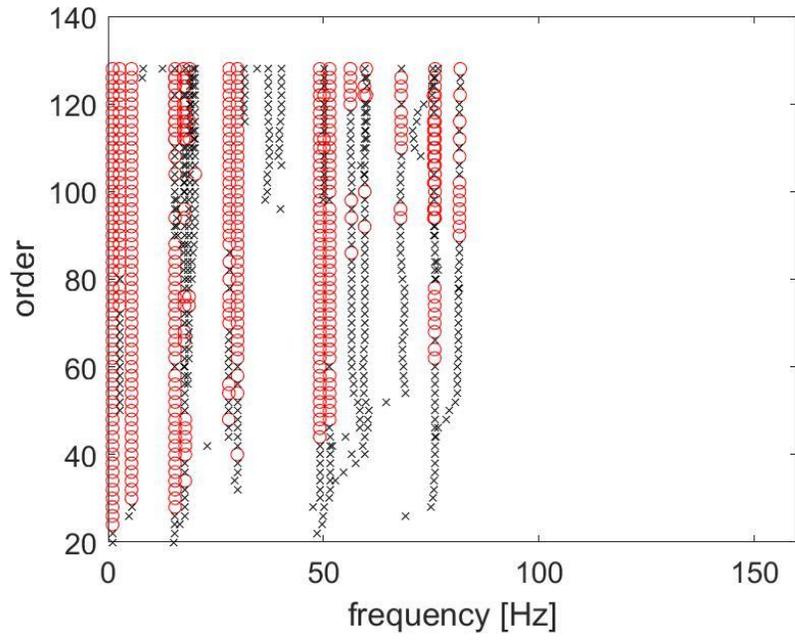
In this step 1141 certainty mathematical and 894 possible physical poles are identified. The stabilization diagrams obtained by the SSI and after applying the HVC are depicted in the following figures.





### Helicopter Blade: Soft Validation Criteria

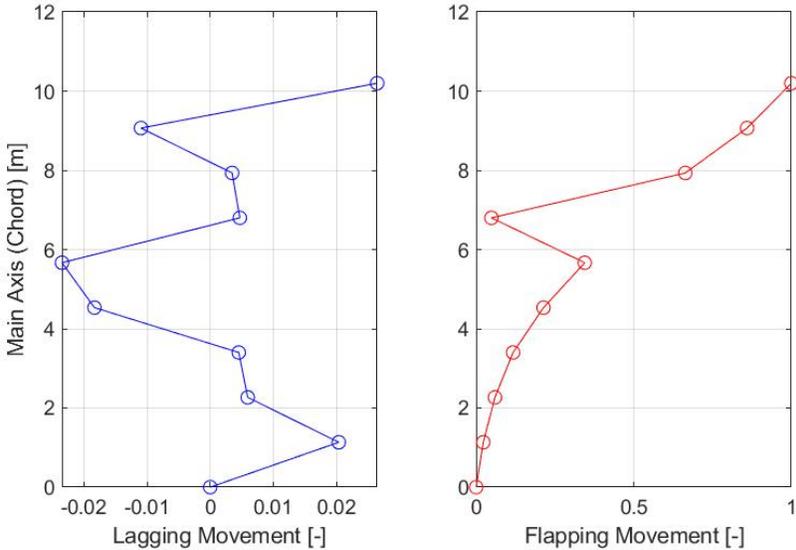
Following the paragraph “Soft Validation Criteria” in **Methodology**, the SVC is applied to the possible physical poles identified by the HVC. In the following figures, the red dots are the stable poles that are successively evaluated during the clustering process, instead the unstable poles are represented as black cross points.



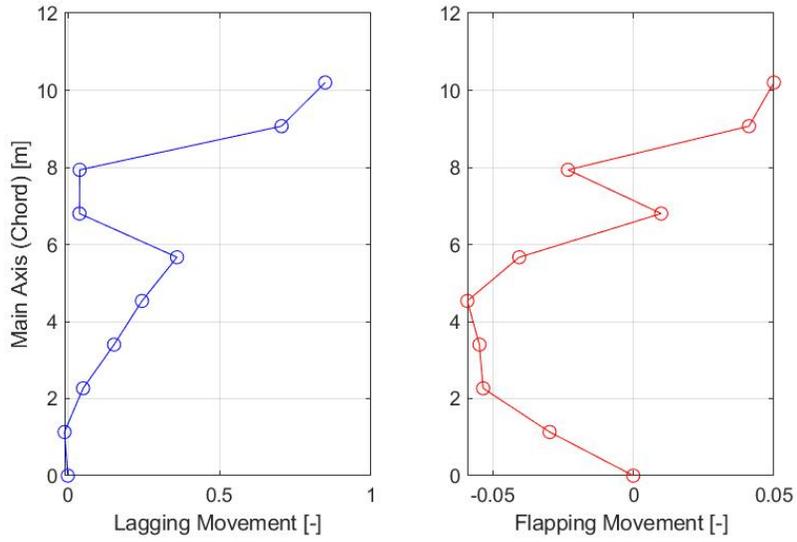
# Helicopter Blade: Mode shapes

The mode shapes identified by the AOMA method are represented in the following figures. Moreover, the type of the mode, i.e. bending or torsional case, is calculated as in **Numerical case** in the paragraph **Final modal parameters**.

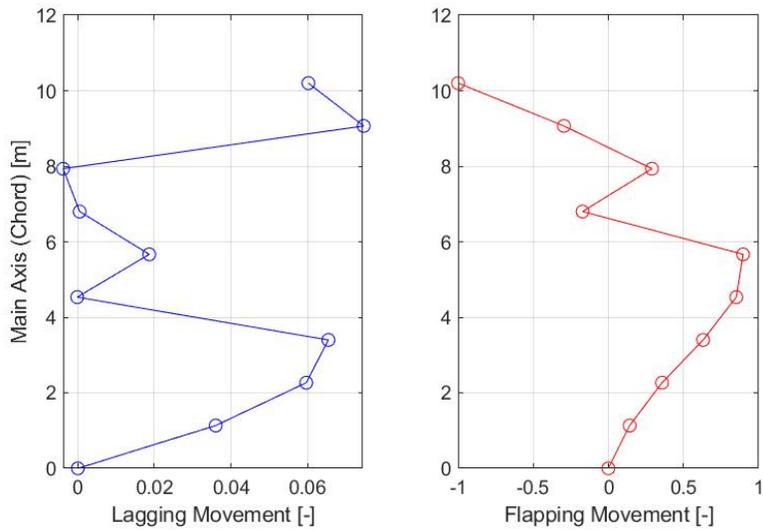
Mode shape 1  
1th Flapping case  
 $f_n$ : 1.0169Hz dr: 0.4564%



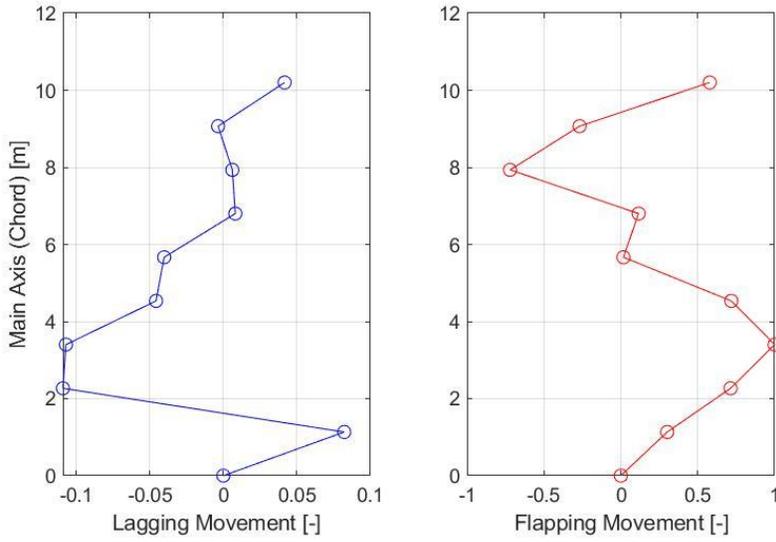
Mode shape 2  
 1th Lagging case  
 $f_n$  : 2.6625Hz dr: 2.8968%



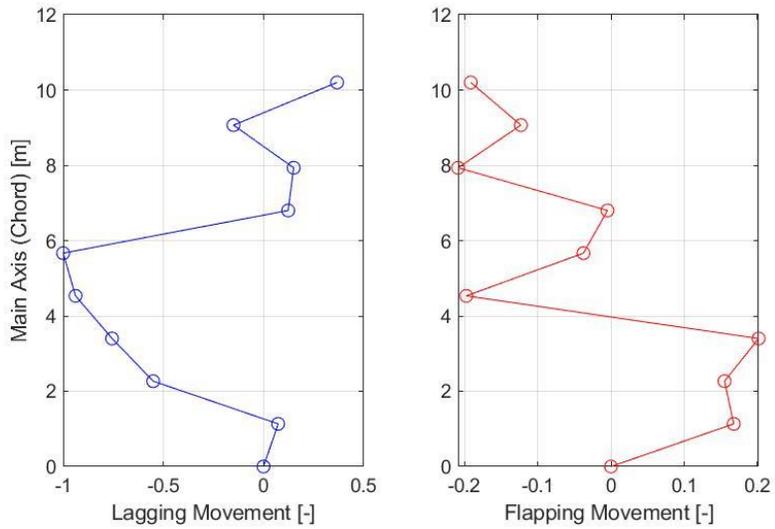
Mode shape 3  
 2th Flapping case  
 $f_n$  : 5.4324Hz dr: 0.5013%



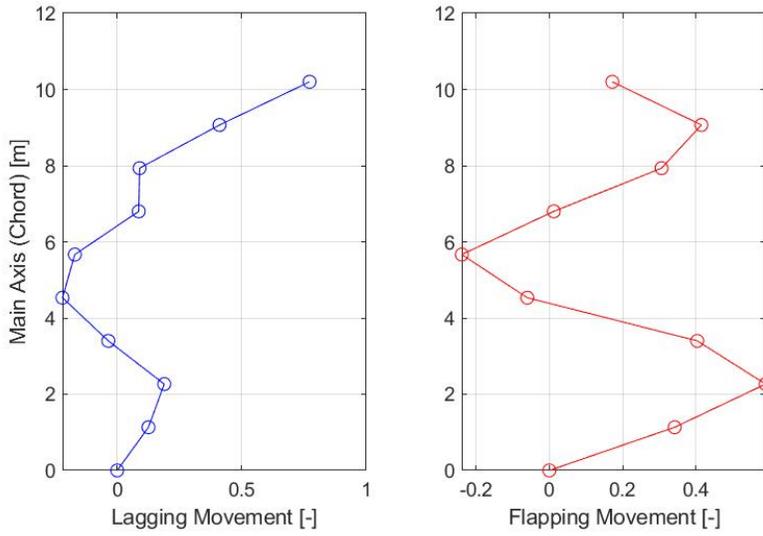
Mode shape 4  
 3th Flapping case  
 $f_n$  : 15.5572Hz dr: 0.1982%



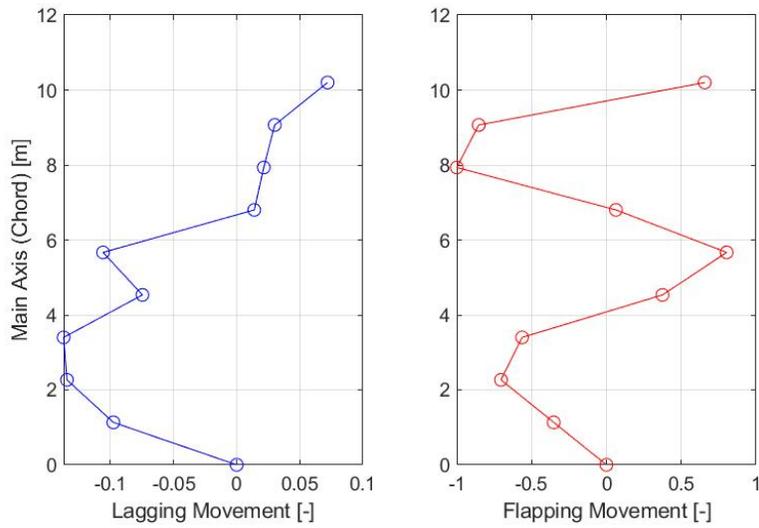
Mode shape 5  
 2th Lagging case  
 $f_n$  : 17.9519Hz dr: 0.7800%



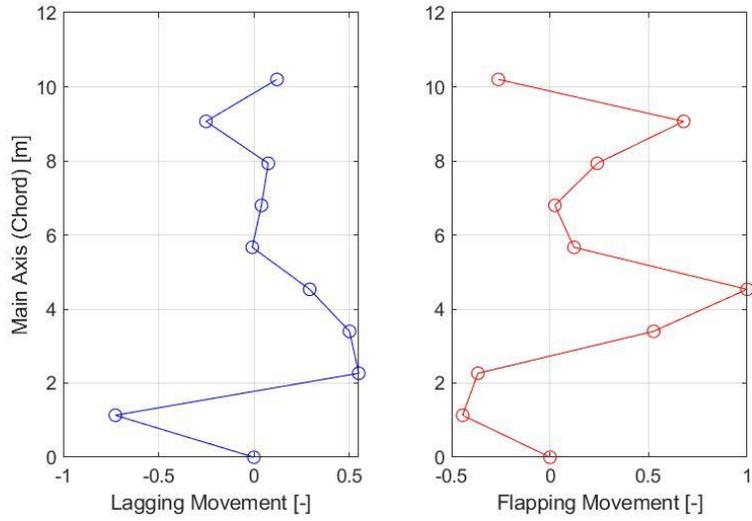
Mode shape 6  
 1th Torsional case  
 $f_n$  : 28.1241Hz dr: 0.6773



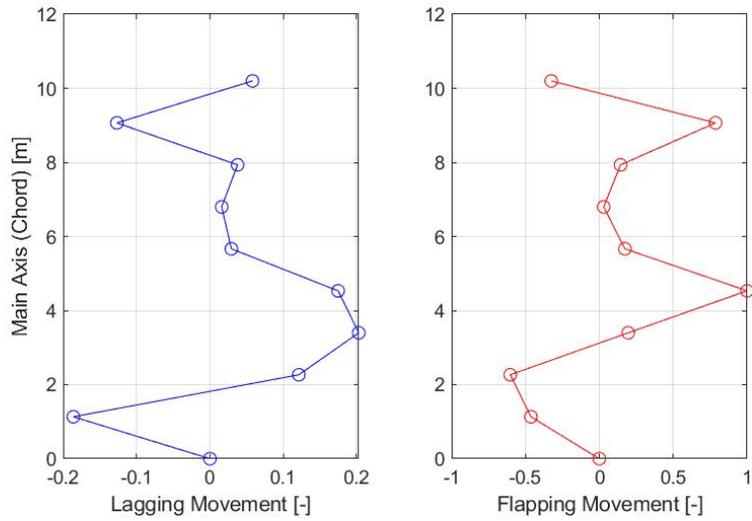
Mode shape 7  
 4th Flapping case  
 $f_n$  : 30.0562Hz dr: 0.3865%



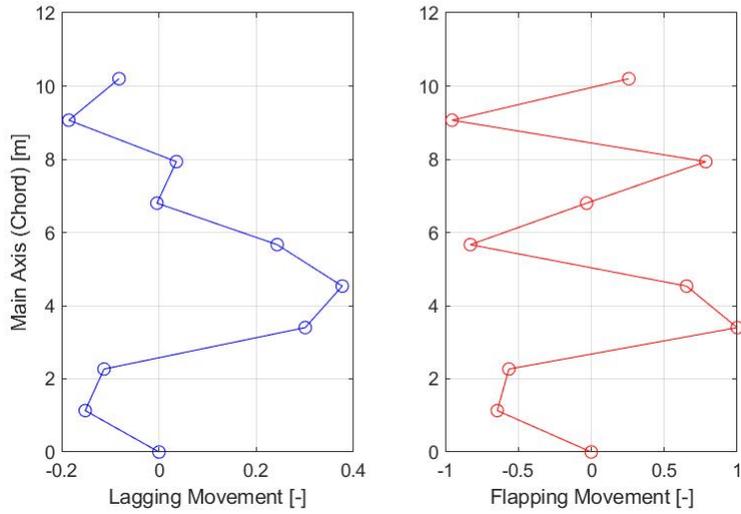
Mode shape 8  
 1th Spurious case  
 $f_n$ : 49.1232Hz dr: 0.0159%



Mode shape 9  
 5th Flapping case  
 $f_n$ : 51.4349Hz dr: 0.3600%



Mode shape 10  
 6th Flapping case  
 $f_n$ : 75.7893Hz dr: 0.5117%



Mode shape 11  
 2th Torsional case  
 $f_n$ : 81.6041Hz dr: 1.7679%

